08/824,178

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C12N 15/11, C12Q 1/68 — A2

(11) International Publication Number: **WO 93/16178**

(43) International Publication Date: 19 August 1993 (19.08.93)

(21) International Application Number: PCT/US93/01294

(22) International Filing Date: 12 February 1993 (12.02.93)

(30) Priority data:
07/837,195 12 February 1992 (12.02.92) US

(71) Applicant: THE UNITED STATES OF AMERICA, as represented by THE SECRETARY, DEPARTMENT OF HEALTH AND HUMAN SERVICES [US/US]; Washington, DC (US).

(72) Inventors: VENTER, Craig, J. ; 1718 Nordic Hill Circle, Silver Spring, MD 20906 (US). ADAMS, Mark, D. ; 12812 Sage Terrace, Germantown, MD 20874 (US). MORENO, Ruben, F. ; 14415 Coral Gables Way, North Potomac, MD 20878 (US).

(74) Agents: ALTMAN, Daniel, E. et al.; Knobbe, Martens, Olson and Bear, 620 Newport Center Drive, 16th Floor, Newport Beach, CA 92660 (US).

(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: SEQUENCES CHARACTERISTIC OF HUMAN GENE TRANSCRIPTION PRODUCT

(57) Abstract

Partial and complete human cDNA and genomic sequences corresponding to particular expressed sequence tags (ESTs). The ESTs are cDNA sequences that are generally between 150 and 500 base pairs in length, are derived from human brain cDNA libraries, correspond to genes transcribed in human brain, and have base sequences identified herein as SEQ ID NOS: 1-2421.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**FR**	France	**MR**	Mauritania
AU	Australia	**GA**	Gabon	**MW**	Malawi
BB	Barbados	**GB**	United Kingdom	**NL**	Netherlands
BE	Belgium	**GN**	Guinea	**NO**	Norway
BF	Burkina Faso	**GR**	Greece	**NZ**	New Zealand
BG	Bulgaria	**HU**	Hungary	**PL**	Poland
BJ	Benin	**IE**	Ireland	**PT**	Portugal
BR	Brazil	**IT**	Italy	**RO**	Romania
CA	Canada	**JP**	Japan	**RU**	Russian Federation
CF	Central African Republic	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CG	Congo			**SE**	Sweden
CH	Switzerland	**KR**	Republic of Korea	**SK**	Slovak Republic
CI	Côte d'Ivoire	**KZ**	Kazakhstan	**SN**	Senegal
CM	Cameroon	**LI**	Liechtenstein	**SU**	Soviet Union
CS	Czechoslovakia	**LK**	Sri Lanka	**TD**	Chad
CZ	Czech Republic	**LU**	Luxembourg	**TG**	Togo
DE	Germany	**MC**	Monaco	**UA**	Ukraine
DK	Denmark	**MG**	Madagascar	**US**	United States of America
ES	Spain	**ML**	Mali	**VN**	Viet Nam
FI	Finland	**MN**	Mongolia		

SEQUENCES CHARACTERISTIC OF HUMAN GENE TRANSCRIPTION PRODUCT

Technical Field

The present invention relates to newly identified polynucleotide sequences corresponding to transcription products of human genes, and to complete gene sequences associated therewith.

Background

This invention relates to human genes. Identification and sequencing of human genes is a major goal of modern scientific research. The sequence of human genes is more than just a scientific curiosity. For example, by identifying genes and determining their sequences, scientists have been able to make large quantities of valuable human "gene products." These include human insulin, interferon, Factor VIII, tumor necrosis factor, human growth hormone, tissue plasminogen activator, and numerous other compounds. Additionally, knowledge of gene sequences can provide the key to treatment or cure of genetic diseases (such as muscular dystrophy and cystic fibrosis). The present invention represents a quantum leap forward in mankind's knowledge of human gene sequences.

There are several basic concepts of molecular biology which figure prominently in the invention. A brief explanation of those concepts follows. Additional background information and definitions for scientific terms can be found

in the literature. See, for example, "Glossary of Genetics, Classical and Molecular" by R. Rieger, A. Michaelis, and M.M. Green (Fifth Edition, Springer-Verlag, New York (1991)). The contents of this and other publications cited in the specification are incorporated by reference herein.

At an initial level, the present invention is based on identification and characterization of gene segments. Genes are the basic units of inheritance. Each gene is a string of connected bases called nucleotides. Most genes are formed of deoxyribonucleic acid, DNA. (Some viruses contain genes of ribonucleic acid, RNA.) The genetic information resides in the particular sequence in which the bases are arranged. A short sequence of nucleotides is often called a polynucleotide or an oligonucleotide.

Like genes, polypeptides are built from long strings of individual units. These units are amino acids. The nucleotide sequence of a gene tells the cell the sequence in which to arrange the amino acids to make the polypeptide encoded by that gene. In general, chains of up to about 200 amino acids are called polypeptides, while proteins are larger molecules made up of polypeptide subunits; both types of molecules are referred to generally herein as polypeptides. A triplet of nucleotides (codon) in DNA codes for each amino acid or signals the beginning or end of the message (anticodon). The term codon is also used for the corresponding (and complementary) sequences of three nucleotides in the mRNA into which the original DNA sequence is transcribed.

Generally, enzymes in the cell transcribe the permanent DNA of the gene into a temporary RNA copy, called messenger RNA or mRNA. The mRNA, in turn, can be translated into a polypeptide by the cell. This entire process is called gene expression, and the polypeptide is the gene product encoded by the gene.

Scientists have previously discovered how to reverse the transcription process and copy mRNA back into DNA using an

enzyme called reverse transcriptase. The resulting is called complementary DNA, or cDNA. This is schematically shown in the single Figure. When substantially all of the mRNA from one cell or tissue is converted to cDNA at once and cloned into multiple copies of a recombinant vector to allow replication and manipulation in the laboratory, the result is called a cDNA library.

The various types of genes include those which code for polypeptides, those which are transcribed into RNA but are not translated into polypeptides, and those whose functional significance does not demand that they be transcribed at all. Most genes are found on large molecules of DNA located in chromosomes. Double stranded cDNA carries all the information of a gene. Each base of the first strand is joined to a complementary base (hybridized) in the second strand. The linear DNA molecules in chromosomes have thousands of genes distributed along their length. Chromosomes include both coding regions (coding for polypeptides) and noncoding regions; the coding regions represent only about three percent of the total chromosome sequence.

An individual gene has regulatory regions that include a promoter which directs expression of the gene, a coding region which can code for a polypeptide, and a termination signal. The regulatory DNA sequence is usually a noncoding region that determines if, where, when, and at what level a particular gene is expressed.

The coding regions of many genes are discontinuous, with coding sequences (exons) alternating with noncoding regions (introns). The final mRNA copy of the gene does not include these introns (which can be much longer than the coding region itself), although it does contain certain untranslated regions that usually do not code for the polynucleotide gene product. Untranslated sequences at the beginning and end of the mRNA are known as 5'- and 3'-untranslated regions,

respectively. This nomenclature reflects the orientation of the nucleotide constituents of the mRNA.

A cDNA is a DNA copy of a messenger RNA, which contains all of the exons of a gene. The cDNA can be thought of as having three parts: an untranslated 5' leader, an uninterrupted polypeptide-coding sequence, and a 3' untranslated region. The untranslated leader and trailing sequences are important for initiation of translation, mRNA stability, and other functions. The untranslated leader and trailing sequences are called 5'- and 3'-untranslated sequences, respectively. The 3' untranslated sequence is usually longer than the 5' untranslated leader, and can be longer than the polypeptide-coding sequence. The untranslated regions typically have many, randomly-distributed stop codons, and do not display the nonrandom base arrangements found in coding sequences. The 5'-untranslated sequence is relatively short, generally between 20 and 200 bases. The 3'-untranslated sequence is often many times longer, up to several thousand bases.

The translated or coding sequence begins with a translational start codon (AUG or GUG) and ends with a translational stop codon (UAA, UGA, or UAG). Generally, translation begins at the first "start" codon on the mRNA and proceeds to the first "stop" codon. Coding sequences can be distinguished by their nonrandom distribution of bases; numerous computer algorithms have been developed to distinguish coding from noncoding regions in this way.

Human DNA differs from person to person. No two persons (except perhaps identical twins) have identical DNA. While the differences, called allelic variations or polymorphisms, are slight on a molecular level, they account for most of the physical and other observable differences between individuals. It has been estimated that approximately 14 million sequence polymorphism differences exist between individuals.

The ability of one strand of DNA to attach or hybridize to a complementary strand has already been exploited for several purposes. For example, small pieces of DNA (15 to 25 base pairs long) can be made which will hybridize to longer strands of DNA which have a complementary sequence. These short "primers" can be selected such that they hybridize to a specific, unique location on the longer strand. Once the primers have hybridized to their target on the DNA, the polymerase chain reaction (PCR) can be employed to generate millions of copies of (or amplify) the particular segment of DNA between the locations to which two primers are bound. Briefly, this technique allows amplification of a DNA region situated between two convergent primers, using oligonucleotide primers that hybridize to opposite strands. Primer extension proceeds inward across the region between the two primers, and the product of DNA synthesis of one primer serves as a template for the other primer. Repeated cycles of DNA denaturation, annealing of primers, and extension result in an exponential increase in the number of copies of the region bounded by the primers.

Similarly, a labeled segment of single-stranded DNA can be hybridized to a longer DNA sequence, such as a chromosome, to mark a specific location on the longer sequence. Segments of DNA 50 bases long or longer that hybridize to a unique DNA location in the human genome are extremely unlikely to hybridize elsewhere in the human genome.

The Human Genome Project is an effort to sequence all human DNA (the human genome). The human genome is estimated to comprise 50,000 - 100,000 genes, up to 30,000 of which might be expressed in the brain (Sutcliffe, **Ann. Rev. Neurosci. 11**:157 (1988)). Once dedicated human chromosome sequencing begins in three to five years, it was expected that 12-15 years will be required to complete the sequence of the genome (Report of the Ad Hoc Program Advisory Committee on Complex Genomes, Reston, Va., Feb. 1988, D. Baltimore Ed. (NIH, Bethesda, Md, 1988)). At that rate, the majority of

human genes would remain unknown for at least the next decade. The present invention can greatly accelerate the pace at which human genes can be identified and mapped. Most gene researchers, in conjunction with publication of their results in this field, submit sequence data to the GenBank database. Prior to the present invention, GenBank listed the sequences of only a few thousand human genes and less than two hundred human brain mRNAs (GenBank Release 66.0, December, 1990).

The role of sequencing complementary DNA (cDNA), reverse transcribed from mRNA, as a part of the human genome project has been vigorously debated since the idea of determining the complete nucleotide sequence of humans first surfaced. The coding sequence of all human genes represents most of the information content of the genome, but only 3-5% of the total DNA. In contrast, cDNA (which is only made from the transcription product of active genes) is one-half to three-fourths (the remainder being 5'- and 3'-untranslated sequence) meaningful genetic information. Thus, some have argued that cDNA sequencing should take precedence over genomic sequencing (Brenner, **CIBA Found. Symp. 149**:6 (1990)). However, until now, such arguments have not been heeded.

Genomic sequencing proponents have argued the difficulty of finding every mRNA expressed in all tissues, cell types, and developmental states, and that much valuable information from intronic and intergenic regions, including control and regulatory sequences, will be missed by cDNA sequencing. (Report of the Committee on Mapping and Sequencing the Human Genome, National Research Council (National Academy Press, Washington, D.C. 1988)). Further, sequencing of transcribed regions of the genome using cDNA libraries has heretofore been considered impractical or unsatisfactory. Libraries of cDNA were believed to be dominated by repetitive elements, mitochondrial genes, ribosomal RNA genes, and other nuclear genes comprising common or housekeeping sequences. It was believed that cDNA libraries would provide few sequences

corresponding to structural and regulatory polypeptides or peptides. See, for example, Putney, et al., **Nature 302:718-721 (1983)**. Putney, et al. sequenced over 150 clones from a rabbit muscle cDNA library and identified clones for 13 of the 19 known muscle polypeptides, including one new isotype but no unknown coding sequences.

Another perceived drawback of cDNA sequencing was that some mRNAs are abundant, and some are rare. The cellular quantities of mRNA from various genes can vary by several orders of magnitude. This led critics to believe that most information obtained from cDNA sequencing would be repetitious and useless.

The present invention demonstrates that, despite such skepticism, cDNA sequencing now provides a rapid method for obtaining enormous amounts of valuable genetic information and DNA products of great utility for the biotechnology and pharmaceutical industries. Not only can many distinct cDNAs be isolated and sequenced, even partial cDNAs can be used, with conventional, well-understood methods, to isolate entire genes, and to determine the chromosomal locations and biological functions of these genes. As is demonstrated here, fragments of only a few hundred bases are sufficient, in many cases, to identify the probable function of a new human gene if it is similar in structure to a gene from another animal, or from plants or bacteria. Similarly, even fragments of untranslated regions of a cDNA can be used to: i) isolate the coding sequence of the cDNA; ii) isolate the complete gene; iii) determine the position of the gene on a human chromosome, and hence the potential of the gene to cause a human genetic disease; and iv) determine the function of the gene by means of experiments in which the function of the native gene is disrupted by the addition of a short DNA fragment to the cell, e.g., using triple helix or antisense probes.

Because coding regions comprise such a small portion of the human genome, identification and mapping of transcribed

regions and coding regions of chromosomes is of significant interest. There is a corresponding need for reagents for identifying and marking coding regions and transcribed regions of chromosomes. Furthermore, such human sequences are valuable for chromosome mapping, human identification, identification of tissue type and origin, forensic identification, and locating disease-associated genes (i.e., genes that are associated with an inherited human disease, whether through mutation, deletion, or faulty gene expression) on the chromosome.

SUMMARY OF THE INVENTION

Contrary to the expectations of the scientific community, cDNA screening and sequencing techniques have now been used to discover a large number of heretofore unknown human genes. Disclosed herein are over 2,400 new human polynucleotide sequences. These sequences could represent up to 5% of all human genes. The novelty of these sequences has been established through comparison to both nucleotide sequence databases and amino acid sequence databases. Surprisingly, over 80% of the sequences generated were unrelated to any sequences previously described in the literature.

The sequences of the present invention were ascertained using a fast approach to cDNA characterization. This approach could facilitate the tagging of most expressed human genes within a few years at a fraction of the cost of complete genomic sequencing, provide new genetic markers, provide new DNA-based therapeutics and diagnostics, and provide other valuable nucleotide reagents.

The sequences disclosed herein, styled Expressed Sequence Tags ("ESTs"), are markers for human genes actually transcribed *in vivo*. Techniques are disclosed for using these ESTs to obtain the full coding region of the corresponding gene. The use of ESTs, complete coding sequences, or fragments thereof for marking chromosomes, for

mapping locations of expressed genes on chromosomes, for individual or forensic identification, for mapping locations of disease-associated genes, for identification of tissue type, and for preparation of antisense sequences, probes, and constructs is discussed in detail below. Unlike the random genomic DNA sequence tagged sites (STSs) (Olson et al., **Science 245**:1434 (1989)), ESTs point directly to expressed genes.

Various aspects of the present invention thus include the individual ESTs, corresponding partial and complete cDNA, genomic DNA, mRNA, antisense strands, triple helix probes, PCR primers, coding regions, and constructs. Also, where one skilled in the art is enabled by this specification to prepare expression vectors and polypeptide expression products, they are also within the scope of the present invention, along with antibodies, especially monoclonal antibodies, to such expression products.

BRIEF DESCRIPTION OF THE DRAWING

The single drawing Figure schematically illustrates the progression from chromosome to gene to mRNA to cDNA.

DETAILED DESCRIPTION OF THE INVENTION

The detailed description that follows provides not only the actual sequence of each new EST, but also explains how the ESTs were obtained, how to obtain the corresponding complete cDNA sequence and the corresponding genomic DNA sequence, how to make DNA constructs from the ESTs and corresponding sequences, how to use those sequences as reagents in molecular biology and other fields, how to produce gene products from the ESTs and corresponding sequences and antibodies to those gene products, and the functional categories of many ESTs and corresponding genes. Furthermore, numerous actual working examples and predictive

examples are provided to demonstrate and exemplify numerous aspects of the invention.

I. **ESTs from cDNA Libraries**

The sequences of the present invention were isolated from commercially available and custom made cDNA libraries using a rapid screening and sequencing technique. In general, the method comprises applying conventional automated DNA sequencing technology to screening clones, advantageously randomly selected clones, from a cDNA library. Preferably, the library is initially "enriched" through removal of ribosomal sequences and other common sequences prior to clone selection. According to the present method, ESTs are generated from partial DNA sequencing of the selected clones. The ESTs of the present invention were generated using low redundancy of sequencing, typically a single sequencing reaction. While single sequencing reactions may have an accuracy as low as 97%, this nevertheless provides sufficient fidelity for identification of the sequence and design of PCR primers.

Most human genes can be identified by EST sequencing from libraries of cDNA copies of messenger RNAs. However, some genes are expressed only at specific times during embryonic development, or only in small amounts in a few specific cell types. Other genes have mRNAs that are degraded very quickly by the cell in which they are expressed. If any of these are the case, transcripts of the gene will not be represented in cDNA libraries so the gene will not be identifiable by EST sequencing. A new method called "exon amplification", however, can be used to isolate and identify transcripts of such genes.

Exon amplification works by artificially expressing part or all of a gene that is contained in a cloned fragment of genomic DNA such as a cosmid or yeast artificial chromosome (YAC). The gene is cloned into a special vector, designed at MIT, that uses control elements from virus genes to express

the protein-coding exons of the human gene of interest. Exon trapping shows considerable promise as a general technique for identifying those genes in the human genome that cannot be found by cDNA cloning and EST sequencing. Exon amplification will also be useful for identifying the genes in regions of genomic DNA to which disease genes have been mapped. The exon amplification method can be used directly with the cosmid and YAC clones frown human chromosomes that are being obtained by both NIH and DOE supported human genome centers. ESTs comprise DNA sequences corresponding to a portion of nuclear encoded messenger RNA. An EST is of sufficient length to permit: (1) amplification of the specific sequence from a cDNA library, e.g., by polymerase chain reaction (PCR); (2) use of a synthetic polynucleotide corresponding to a partial or complete sequence of the EST as a hybridization probe of a cDNA library, generally having 30 - 50 base pairs; or (3) unique designation of the pure cDNA clone from which the EST was derived (the EST clone) for use as a hybridization probe of a cDNA library. Preferably, EST-derived primer pairs and sequences amplify or detectably hybridize to a sequence from a genomic library.

It has been found that sufficient information is contained in the 150-400 base ESTs from one sequencing run to effect preliminary identification and exact chromosome mapping. Accordingly, the ESTs disclosed herein are generally at least 150 base pairs in length. The length of an EST is determined by the quality of sequencing data and the length of the cloned cDNA. Raw data from the automated sequencers is edited to remove low quality sequence at the end of the sequencing run. High quality sequences (usually a result of sequencing templates without excessive salt contamination) generally give about 400 bp of reliable sequence data; other sequences give fewer bases of reliable data. A 150 bp EST is long enough to be translated into a 50 amino acid peptide sequence. This length is sufficient to observe similarities when they exist in a database search. Furthermore, 150 bp is

long enough to design PCR primers from each end of the sequence to amplify the complete EST. Sequences shorter than 150 bp are difficult to purify and use following PCR amplification. Furthermore, a 150 bp polynucleotide is likely to give a very strong signal with low background in a screen of a genomic library.

Finally, it is highly unlikely that a sequence of the same 150 bp exists in any genes in the genome besides the one tagged by the EST. Some closely related gene family members have very similar nucleotide sequences, but no examples of pairs of human genes with long segments of identical sequence have been reported to date. For instance, there are three known β-tubulin genes in humans. Several ESTs were found that matched one or another of these tubulin genes, but several new members of this gene family were also found and could be clearly distinguished from the three known members. ESTs that match perfectly to several different genes can be detected by hybridizing to chromosomes: if many chromosomal loci are observed, the sequence (or a close variant) is present in more than one gene. This problem can be circumvented by using the 3'-untranslated part of the cDNA alone as a probe for the chromosomal location or for the full-length cDNA or gene. The 3'-untranslated region is more likely to be unique within gene families, since there is no evolutionary pressure to conserve a coding function of this region of the mRNA.

As demonstrated in the Examples that follow, ESTs can be used to map the expressed sequence to a particular chromosome. In addition, ESTs can be expanded to provide the full coding regions, as detailed below. In this manner, previously unknown genes can be identified.

While a variety of cDNA libraries can be used to obtain ESTs, human brain cDNA libraries are exemplified and represent a preferred embodiment. Suitable cDNA libraries can be freshly prepared or obtained commercially, e.g., as shown in Examples 1, 2, and 11. The cDNA libraries from the

desired tissue are preferably preprocessed by conventional techniques to reduce repeated sequencing of high and intermediate abundance clones and to maximize the chances of finding rare messages from specific cell populations. Preferably, preprocessing includes the use of defined composition prescreening probes, e.g., cDNA corresponding to mitochondria, abundant sequences, ribosomes, actins, myelin basic polypeptides, or any other known high abundance peptide; these prescreening probes used for preprocessing are generally derived from known ESTs. Other useful preprocessing techniques include subtraction, which preferentially reduces the population of certain sequences in the library (e.g., see A. Swaroop et al., **Nucl. Acids Res.** 19, 1954 (1991)), and normalization, which results in all sequences being represented in approximately equal proportions in the library (Patanjali et al, **Proc. Natl. Acad. Sci. USA 88**:1943 (1991)).

The cDNA libraries used in the present method will ideally use directional cloning methods so that either the 5' end of the cDNA (likely to contain coding sequence) or the 3' end (likely to be a non-coding sequence) can be selectively obtained."

Libraries of cDNA can also be generated from recombinant expression of genomic DNA. After they are amplified, ESTs can be obtained and sequenced, e.g., as illustrated in Example 11.

The sequences of the present invention include the specific sequences set forth in the Sequence Listing and designated SEQ ID NO: 1 - SEQ ID NO: 2412. In one aspect of this embodiment, the invention relates to those sequences of SEQ ID NOS: 1 - 2412 that comprise the cDNA coding sequences for polypeptides having less than 95% identity with known amino acid sequences (see Table 2) and more preferably less than 90% or 85% identity. In a second aspect, the invention relates to those sequences of SEQ ID NOS: 1 - 2412 that encode polypeptides having no similarity to known amino acid

sequences (see Examples that follow). Precisely because they do not contain coding regions and are therefore more unique in their sequence structures, those sequences which meet neither of the preceding criteria can be most useful and are generally preferred for mapping.

Consistent with the NIH mission and its responsibilities to disseminate knowledge and share the tangible fruits of its research, the present inventors have taken a number of steps to facilitate sequence data and clone availability. All EST sequences have been submitted to GenBank (representing an addition equivalent to 7% of the human nucleotides in Release 69 of GenBank, September 1991). The corresponding cDNA clones have been submitted to the American Type Culture Collection and information on clones and sequences has been submitted to the Genome Data Base (Pearson, **P. Nucl. Acids Res. 19 (Suppl.)**: 2237-9 (1991)).

II. <u>Complete Coding Sequences from ESTs</u>

The ESTs of the present invention generally represent relatively small coding regions or untranslated regions of human genes. Although most of these sequences do not code for a complete gene product, the ESTs of the present invention are highly specific markers for the corresponding complete coding regions. The ESTs are of sufficient length that they will hybridize, under stringent conditions, only with DNA for that gene to which they correspond. Suitably stringent conditions comprise conditions, for example, where at least 95%, preferably at least 97% or 98% identity (base pairing), is required for hybridization. This property permits use of the EST to isolate the entire coding region and even the entire sequence. Therefore, only routine laboratory work is necessary to parlay the unique EST sequence into the corresponding unique complete gene sequence.

Thus, each of the ESTs of the present invention "corresponds" to a particular unique human gene. Knowledge

of the EST sequence permits routine isolation and sequencing of the complete coding sequence of the corresponding gene. The complete coding sequence is present in a full-length cDNA clone as well as in the gene carried on genomic clones. Therefore, each EST "corresponds" to a cDNA (from which the EST was derived), a complete genomic gene sequence, a polypeptide coding region (which can be obtained either from the cDNA or genomic DNA), and a polypeptide or amino acid sequence encoded by that region.

The first step in determining where an EST is located in the cDNA is to analyze the EST for the presence of coding sequence, e.g., as described in Example 14. The CRM program predicts the extent and orientation of the coding region of a sequence. Based on this information, one can infer the presence of start or stop codons within a sequence and whether the sequence is completely coding or completely non-coding. If start or stop codons are present, then the EST can cover both part of the 5'-untranslated or 3'-untranslated part of the mRNA (respectively) as well as part of the coding sequence. If no coding sequence is present, it is likely that the EST is derived from the 3'-untranslated sequence due to its longer length and the fact that most cDNA library construction methods are biased toward the 3' end of the mRNA.

One general procedure for obtaining complete sequences from ESTs is as follows:

1. Purify selected human DNA from an EST clone (the cDNA clone that was sequenced to give the EST), e.g., by endonuclease digestion using ECOR1, gel electrophoresis, and isolation of the aforementioned clone by removal from low-melting agarose gel.

2. Radiolabel the isolated insert DNA, e.g., with ^{32}P labels, preferably by nick translation or random primer labeling.

3. Use the labeled EST insert as a probe to screen a lambda phage cDNA library or a plasmid cDNA library.

4. Identify colonies containing clones related to the probe cDNA and purify them by known purification methods.

5. Nucleotide sequence the ends of the newly purified clones to identify full length sequences.

6. Perform complete sequencing of full length clones by Exonuclease III digestion or primer walking. Northern blots of the mRNA from various tissues using at least part of the EST clone as a probe can optionally be performed to check the size of the mRNA against that of the purported full length cDNA.

An EST is a specific tag for a messenger RNA molecule. The complete sequence of that messenger RNA, in the form of cDNA, can be determined using the EST as a probe to identify a cDNA clone corresponding to a full-length transcript, followed by sequencing of that clone. The EST or the full-length cDNA clone can also be used as a probe to identify a genomic clone or clones that contain the complete gene including regulatory and promoter regions, exons, and introns.

ESTs are used as probes to identify the cDNA clones from which an EST was derived. ESTs, or portions thereof, can be nick-translated or end-labelled with P^{32} using polynucleotide kinase using labelling methods known to those with skill in the art (**Basic Methods in Molecular Biology**, L.G. Davis, M.D. Dibner, and J.F. Battey, ed., Elsevier Press, NY, 1986). The lambda library can be directly screened with the labelled ESTs of interest or the library can be converted en masse to pBluescript (Stratagene, La Jolla, California) to facilitate bacterial colony screening. Both methods are well known in the art. Briefly, filters with bacterial colonies containing the library in pBluescript or bacterial lawns containing lambda plaques are denatured and the DNA is fixed to the filters. The filters are hybridized with the labelled probe using hybridization conditions described by Davis et al. The ESTs, cloned into lambda or pBluescript, can be used as positive controls to assess background binding and to adjust

the hybridization and washing stringencies necessary for accurate clone identification. The resulting autoradiograms are compared to duplicate plates of colonies or plaques; each exposed spot corresponds to a positive colony or plaque. The colonies or plaques are selected, expanded and the DNA is isolated from the colonies for further analysis and sequencing.

The ESTs can additionally be used to screen Northern blots of mRNA obtained from various tissues or cell cultures, including the tissue of origin of the EST clone. Northern analysis will most often produce one to several positive bands. The bands can be selected for further study based on the predicted size of the mRNA.

Positive cDNA clones in phage lambda are analyzed to determine the amount of additional sequence they contain using PCR with one primer from the EST and the other primer from the vector. Clones with a larger vector-insert PCR product than the original EST clone are analyzed by restriction digestion and DNA sequencing to determine whether they contain an insert of the same size or similar as the mRNA size on a Northern blot.

Once one or more overlapping cDNA clones are identified, the complete sequence of the clones can be determined. The preferred method is to use exonuclease III digestion (McCombie, W.R, Kirkness, E., Fleming, J.T., Kerlavage, A.R., Iovannisci, D.M., and Martin-Gallardo, R., ***Methods:*** 3: 33-40, 1991). A series of deletion clones is generated, each of which is sequenced. The resulting overlapping sequences are assembled into a single contiguous sequence of high redundancy (usually three to five overlapping sequences at each nucleotide position), resulting in a highly accurate final sequence.

A similar screening and clone selection approach can be applied to obtaining cosmid or lambda clones from a genomic DNA library that contains the complete gene from which the EST was derived (Kirkness, E.F., Kusiak, J.W., Menninger, J.,

Gocayne, J.D., Ward, D.C., and Venter, J.C. *Genomics* 10: 985-995 (1991). Although the process is much more laborious, these genomic clones can be sequenced in their entirety also. A shotgun approach is preferred to sequencing clones with inserts longer than 10 kb (genomic cosmid and lambda clones). In shotgun sequencing, the clone is randomly broken into many small pieces, each of which is partially sequenced. The sequence fragments are then aligned to produce the final contiguous sequence with high redundancy. An intermediate approach is to sequence just the promoter region and the intron-exon boundaries and to estimate the size of the introns by restriction endonuclease digestion (ibid.).

Using the sequence information provided herein, the polynucleotides of the present invention can be derived from natural sources or synthesized using known methods. The sequences falling within the scope of the present invention are not limited to the specific sequences described, but include human allelic and species variations thereof and portions thereof of at least 15-18 bases. (Sequences of at least 15-18 bases can be used, for example, as PCR primers or as DNA probes.) In addition, the invention includes the entire coding sequence associated with the specific polynucleotide sequence of bases described in the Sequence Listing, as well as portions of the entire coding sequence of at least 15-18 bases and allelic and species variations thereof. Furthermore, to accommodate codon variability, the invention includes sequences coding for the same amino acid sequences as do the specific sequences disclosed herein. Finally, although the error rate in the automated sequencing used in the present invention is small, there remains some chance of error. Therefore, claims to particular sequences should not be so narrowly construed as to require inclusion of erroneously identified bases or to exclude corrections.

Any specific sequence disclosed herein can be readily screened for errors by resequencing each EST in both directions (i.e., sequence both strands of cDNA).

The sequences, constructs, vectors, clones, and other materials comprising the present invention can advantageously be in enriched or isolated form. As used herein, "enriched" means that the concentration of the material is at least about 2, 5, 10, 100, or 1000 times its natural concentration (for example), advantageously 0.01%, by weight, preferably at least about 0.1% by weight. Enriched preparations of about 0.5%, 1%, 5%, 10%, and 20% by weight are also contemplated. Further, removal of clones corresponding to ribosomal RNA and "housekeeping" genes and clones without human cDNA inserts results in a library that is "enriched" in the desired clones.

The term "isolated" requires that the material be removed from its original environment (e.g., the natural environment if it is naturally occurring). For example, a naturally-occurring polynucleotide present in a living animal is not isolated, but the same polynucleotide, separated from some or all of the coexisting materials in the natural system, is isolated.

It is also advantageous that the sequences be in purified form. The term "purified" does not require absolute purity; rather, it is intended as a relative definition. Individual EST clones isolated from a cDNA library have been conventionally purified to electrophoretic homogeneity. The sequences obtained from these clones could not be obtained directly either from the library or from total human DNA. The cDNA clones are not naturally occurring as such, but rather are obtained via manipulation of a partially purified naturally occurring substance (messenger RNA). The conversion of mRNA into a cDNA library involves the creation of a synthetic substance (cDNA) and pure individual cDNA clones can be isolated from the synthetic library by clonal selection. Thus, creating a cDNA library from messenger RNA and subsequently isolating individual clones from that library results in an approximately 10^6-fold purification of the native message. Purification of starting material or

natural material to at least one order of magnitude, preferably two or three orders, and more preferably four or five orders of magnitude is expressly contemplated.

In a cDNA library there are many species of mRNA represented. Each cDNA clone can be interesting in its own right, but must be isolated from the library before further experimentation can be completed. In order to sequence any specific cDNA, it must be removed and separated (i.e. isolated and purified) from all the other sequences. This can be accomplished by many techniques known to those of skill in the art. These procedures normally involve identification of a bacterial colony containing the cDNA of interest and further amplification of that bacteria. Once a cDNA is separated from the mixed clone library, it can be used as a template for further procedures such as nucleotide sequencing.

Although claims to large numbers of ESTs and corresponding sequences are presented herein, the invention is not limited to these particular groupings of sequences. Thus, individual sequences are considered as applicants' discoveries or inventions, as are subgroupings of sequences. All of the functional subgroupings set forth in the tables define groupings for which separate claims are contemplated as being within the scope of this invention. Moreover, in addition to claims to individual clones, it is intended that the present disclosure also support claims to numerical subgroupings. Thus, subgroupings of 50 ESTs (and corresponding sequences) are contemplated (e.g., SEQ ID NOS 1-50, 51-100, 101-150, etc.) as being within the scope of this invention, as are subgroupings of 5, 10, 25, 100, 200, and 500 ESTs and corresponding sequences.

III. DNA Constructs

The present invention also includes recombinant constructs comprising one or more of the sequences as broadly described above. The constructs comprise a vector, such as

a plasmid or viral vector, into which a sequence of the invention has been inserted, in a sense or antisense orientation. In a preferred aspect of this embodiment, the construct further comprises regulatory sequences, including for example, a promoter, operably linked to the sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art, and are commercially available. The following vectors are provided by way of example. **Bacterial:** pBs, phagescript, ϕX174, pBluescript SK, pBs KS, pNH8a, pNH16a, pNH18a, pNH46a (Stratagene); pTrc99A, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia). **Eukaryotic:** pWLneo, pSV2cat, pOG44, pXT1, pSG (Stratagene); pSVK3, pBPV, pMSG, pSVL (Pharmacia).

Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda P_R, and trc. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art.

In a further embodiment, the present invention relates to host cells containing the above-described construct. The host cell can be a higher eukaryotic cell, such as a mammalian cell, or a lower eukaryotic cell, such as a yeast cell, or the host cell can be a procaryotic cell, such as a bacterial cell. Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE dextran mediated transfection, or electroporation (Davis, L., Dibner, M., Battey, I., **Basic Methods in Molecular Biology**, (1986)).

The constructs in host cells can be used in a conventional manner to produce the gene product coded by the recombinant sequence. Alternatively, the encoded polypeptide

can be synthetically produced by conventional peptide synthesizers.

Certain ESTs have already been preliminarily categorized by analogy to related sequences in other organisms (see Table 2). Table 10 of Example 10 categorizes particular ESTs broadly as metabolic, regulatory, and structural sequences where known. Constructs comprising genes or coding sequences corresponding to each of these categories are, therefore, specifically and individually contemplated.

Table 11 more particularly separates 127 new ESTs into 13 categories using a different criteria. These are genes related to cell surface; developmental control; energy metabolism; kinase and phosphatase; oncogenes; other metabolism-related polypeptides; peptidases and peptidase inhibitors; receptors; structural and cytoskeletal; signal transduction; transporters; transcription, translation, and subcellular localization; and transcription factors. Table 11 further identifies the EST by the particular gene product for which it apparently codes. Each of these categories individually comprises a preferred category of EST, and preferred constructs and resulting polypeptide can be prepared from those ESTs or the corresponding complete gene sequence.

IV. ESTs and Corresponding Sequences as Reagents

Each of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. The sequences can be used as diagnostic probes for the presence of a specific mRNA in a particular cell type. In addition, these sequences can be used as diagnostic probes suitable for use in genetic linkage analysis (polymorphisms). Further, the sequences can be used as probes for locating gene regions associated with genetic disease, as explained in more detail below.

The EST and complete gene sequences of the present invention are also valuable for chromosome identification.

Each sequence is specifically targeted to and can hybridize with a particular location on an individual human chromosome. Moreover, there is a current need for identifying particular sites on the chromosome. Few chromosome marking reagents based on actual sequence data (repeat polymorphisms) are presently available for marking chromosomal location. The present invention constitutes a major expansion of available chromosome markers. One hundred ESTS have already been mapped to chromosomes. Using the techniques described in Example 5 or 6, the remaining ESTs and the corresponding complete sequences can similarly be mapped to chromosomes. The mapping of ESTs and cDNAs to chromosomes according to the present invention is an important first step in correlating those sequences with genes associated with disease.

Briefly, sequences can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp) from the ESTs. Computer analysis of the ESTs is used to rapidly select primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers are then used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the EST will yield an amplified fragment.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular EST to a particular chromosome. Three or more clones can be assigned per day using a single thermal cycler. Using the present invention with the same oligonucleotide primers, sublocalization can be achieved with panels of fragments from specific chromosomes or pools of large genomic clones in an analogous manner. Other mapping strategies that can similarly be used to map an EST to its chromosome include <u>in situ</u> hybridization, prescreening with labeled flow-sorted chromosomes and preselection by hybridization to construct chromosome specific cDNA libraries. Results of mapping ESTs to chromosomal segments are listed in Tables 3 and 4.

Fluorescence in situ hybridization (FISH) of a cDNA clone to a metaphase chromosomal spread can be used to provide a precise chromosomal location in one step. This technique can be used with cDNA as short as 500 or 600 bases; however, clones larger than 2,000 bp have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. FISH requires use of the clone from which the EST was derived, and the longer the better. 2,000 bp is good, 4,000 is better, and more than 4,000 is probably not necessary to get good results a reasonable percentage of the time. For a review of this technique, see Verma et al., **Human Chromosomes: a Manual of Basic Techniques.** Pergamon Press, New York (1988).

Reagents for chromosome mapping can be used individually (to mark a single chromosome or a single site on that chromosome) or as panels of reagents (for marking multiple sites and/or multiple chromosomes). Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping (see Tables 8 and 9).

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. (Such data are found, for example, in V. McKusick, **Mendelian Inheritance in Man** (available on line through Johns Hopkins University Welch Medical Library).) The relationship between genes and diseases that have been mapped to the same chromosomal region are then identified through linkage analysis (coinheritance of physically adjacent genes).

Next, it is necessary to determine the differences in the cDNA or genomic sequence between affected and unaffected individuals. If a mutation is observed in some or all of the affected individuals but not in any normal individuals, then

the mutation is likely to be the causative agent of the disease.

With current resolution of physical mapping and genetic mapping techniques, a cDNA precisely localized to a chromosomal region associated with the disease could be one of between 50 and 500 potential causative genes. (This assumes 1 megabase mapping resolution and one gene per 20 kb.)

Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that cDNA sequence. Ultimately, complete sequencing of genes from several individuals is required to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

In addition to the foregoing, the sequences of the invention, as broadly described, can be used to control gene expression through triple helix formation or antisense DNA or RNA, both of which methods are based on binding of a polynucleotide sequence to DNA or RNA. Polynucleotides suitable for use in these methods are usually 20 to 40 bases in length and are designed to be complementary to a region of the gene involved in transcription (triple helix - see Lee et al, **Nucl. Acids Res. 6**: 3073 (1979); Cooney et al, **Science 241**: 456 (1988); and Dervan et al, **Science 251**: 1360 (1991)) or to the mRNA itself (antisense - Okano, **J. Neurochem. 56**: 560 (1991); **Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression**, CRC Press, Boca Raton, FL (1988)). Triple helix formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated to be efficient in model systems. Information contained in the sequences of the present invention is necessary for the design of an antisense or triple helix oligonucleotide.

The present invention is also useful tool in gene therapy, which requires isolation of the disease-associated gene in question as a prerequisite to the insertion of a normal gene into an organism to correct a genetic defect. The high specificity of the cDNA probes according to this invention have promise of targeting such gene locations in a highly accurate manner.

The sequences of the present invention, as broadly defined, are also useful for identification of individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identifying personnel. This method does not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The sequences of the present invention are useful as additional DNA markers for RFLP.

However, RFLP is a pattern based technique, which does not directly focus on the actual DNA sequence of the individual. The sequences of the present invention can be used to provide an alternative technique that determines the actual base-by-base DNA sequence of selected portions of an individual's genome. These sequences can be used to prepare PCR primers for amplifying and isolating such selected DNA. One can, for example, take an EST of the invention and prepare two PCR primers from the 5' and 3' ends of the EST. These are used to amplify an individual's DNA, corresponding to the EST. The amplified DNA is sequenced.

Panels of corresponding DNA sequences from individuals, made this way, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences, due to allelic differences. The sequences of the present invention can be used to particular advantage to

obtain such identification sequences from individuals and from tissue, as explained in Examples 12 - 14.

The EST sequences from Examples 1 and 2 and the complete sequences from Example 13 uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Each of the ESTs or complete coding sequences comprising a part of the present invention can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences of Table 9 for example, could comfortably provide positive individual identification with a panel of perhaps 100 to 1,000 primers which each yield a noncoding amplified sequence of 100 bp. If predicted coding sequences, such as those from Table 6, are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

If a panel of reagents from ESTs or complete sequences of this invention is used to generate a unique ID database for an individual, those same reagents can later be used to identify tissue from that individual. Positive identification of that individual, living or dead can be made from extremely small tissue samples.

Another use for DNA-based identification techniques is in forensic biology. PCR technology can be used to amplify DNA sequences taken from very small biological samples such as tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, semen, etc. In one prior art technique, gene sequences are amplified at specific loci known to contain a large number of allelic variations, for example the DQα class II HLA gene (Erlich, H., **PCR Technology**, Freeman and Co.

(1992)). Once this specific area of the genome is amplified, it is digested with one or more restriction enzymes to yield an identifying set of bands on a Southern blot probed with DNA corresponding to the DQα class II HLA gene.

The sequences of the present invention can be used to provide polynucleotide reagents specifically targeted to additional loci in the human genome, and can enhance the reliability of DNA-based forensic identifications. Those sequences targeted to noncoding regions (see, e.g., Tables 8 and 9) are particularly appropriate. As mentioned above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Reagents for obtaining such sequence information are within the scope of the present invention. Such reagents can comprise complete ESTs or corresponding coding regions, or fragments of either of at least 15 bp, preferably at least 18 bp.

There is also a need for reagents capable of identifying the source of a particular tissue. Such need arises, for example, in forensics when presented with tissue of unknown origin. Appropriate reagents can comprise, for example, DNA probes or primers specific to particular tissue prepared from the ESTs or complete sequences of the present invention. Panels of such reagents can identify tissue by species and/or by organ type. In a similar fashion, these reagents can be used to screen tissue culture for contamination.

V. Production of Polypeptide Corresponding to ESTs

As previously explained, each EST corresponds not only to a coding region, but also to a polypeptide. Once the coding sequence is known, or the gene is cloned which encodes the polypeptide, conventional techniques in molecular biology can be used to obtain the polypeptide.

At the simplest level, the amino acid sequence encoded by the polynucleotide sequence can be synthesized using commercially available peptide synthesizers. This is

particularly useful in producing small peptides and fragments of larger polypeptides. (Fragments are useful, for example, in generating antibodies against the native polypeptide.)

Alternatively, the DNA encoding the desired polypeptide can be inserted into a host organism and expressed. The organism can be a bacterium, yeast, cell line, or multicellular plant or animal. The literature is replete with examples of suitable host organisms and expression techniques. For example, naked polynucleotide (DNA or mRNA) can be injected directly into muscle tissue of mammals, where it is expressed. This methodology can be used to deliver the polypeptide to the animal, or to generate an immune response against a foreign polypeptide. Wolff, et al., **Science 247:1465** (1990); Felgner, et al., **Nature 349:351** (1991). Alternatively, the coding sequence, together with appropriate regulatory regions (i.e., a construct), can be inserted into a vector, which is then used to transfect a cell. The cell (which may or may not be part of a larger organism) then expresses the polypeptide. (See Example 25.)

Antibodies generated against the polypeptide corresponding to a sequence of the present invention can be obtained by direct injection of the naked polypeptide into an animal (as above) or by administering the polypeptide to an animal, preferably a nonhuman. The antibody so obtained will then bind the polypeptide itself. In this manner, even a sequence encoding only a fragment of the polypeptide can be used to generate antibodies binding the whole native polypeptide. Such antibodies can then be used to isolate the polypeptide from tissue expressing that polypeptide. Moreover, a panel of such antibodies, specific to a large number of polypeptides, can be used to identify and differentiate such tissue.

VI. Examples

Certain aspects of the present invention are described in greater detail in the non-limiting Examples that follow.

EXAMPLE 1

cDNA Sequences Determined by Random Clone Selection: First set

METHODOLOGY:

With reference to the data presented in Table 1, lambda ZAP libraries were converted en masse to pBluescript plasmids, transfected into E. coli XL1-Blue cells, and plated on X-gal/IPTG/ampicillin plates. A total of 1058 clones were picked at random from three human brain cDNA libraries: fetal brain, two-year-old hippocampus, and two-year-old temporal cortex (Stratagene catalog #936206, 936205, 935, respectively. Stratagene, 11099 N. Torrey Pines Rd., La Jolla, CA 92037). An analysis of these clones is summarized in Table I (see below) In addition, clones selected from the hippocampus library were also analyzed after subtractive hybridization with the fibroblast library. These results are listed in the "Hippocampus Subtracted" column of Table 1. Templates for DNA sequencing were PCR products or plasmids prepared by the alkaline lysis method. About half of the templates prepared by PCR failed to yield an amplified fragment suitable for sequencing. This was primarily due to use of PCR conditions that minimized the need for further purification of the product but also selected against amplification of long inserts (5 μl fresh or frozen overnight culture of E. coli carrying the pBluescript plasmid, 7.5 μM each dNTP, and 0.1 μM each primer for 35 cycles: 94°C, 40 sec; 55°C, 40 sec; 72°C, 90 sec). A further percentage of the PCR-generated templates failed to sequence, largely due to primer-dimer or other amplification artifacts. Qiagen™ columns improved the percentage of plasmid templates, increasing the yields of usable sequence from about 60% with a standard alkaline lysis protocol to over 90%. Overall, 117 PCR-generated templates and 497 plasmid templates resulted in usable sequence. Dideoxy chain termination sequencing reactions were performed with fluorescent dye-labeled M13

universal or reverse primers. After a cycle sequencing protocol, carried out in a Perkin-Elmer thermal cycler, sequencing reactions were run on an Applied Biosystems, Inc. (Foster City, CA) 373A automated DNA sequencer. (Cycle sequencing was performed in a Perkin Elmer Thermal Cycler for 15 cycles of 95°C, 30 sec; 60°C, 1 sec; 70°C, 60 sec and 15 cycles of 95°C, 30 sec; 70°C, 60 sec with the Applied Biosystems, Inc. Taq Dye Primer Cycle Sequencing Core Kit protocol). Some sequencing reactions were performed on an ABI robotic workstation (Cathcart, **Nature 347**: 310 (1990) hereby incorporated by reference).

RESULTS:

Singe-run DNA sequence data were obtained from 609 randomly chosen cDNA clones. The number of clones sequenced from each library is summarized in Table 1. Double-stranded cDNA clones in the pBluescript vector were sequenced by a cycle sequencing protocol with dye-labeled primers and Applied Biosystems, Inc. 373A DNA Sequences. The average length of usable sequence was 397 bases with a standard deviation of 99 bases.

Subtractive hybridization has been used successfully to reduce the population of highly represented sequences in a cDNA library by selectively removing sequences shared by another library. (Schmid and Girou, **Neurochem. 48**: 307 (1987); Fargnoli et al, **Anal. Biochem. 187**: 364 (1990); Duguid and Dinauer, **Nucl. Acids. Res. 18**: 2789 (1990); Schweinfest, et al, **Genet. Anal. Techn. Appl. 7**: 64 (1990); Travis and Sutcliffe, **Proc. Natl. Acad. Sci. USA 85**: 1696 (1988); Kato, **Eur. J. Neurosci. 2**: 704 (1990)). Subtractive hybridization was therefore tested as a way of enhancing the number of brain-specific clones in the hippocampus library by hybridizing the hippocampus library with a WI38 human lung fibroblast cell line cDNA library and removing the common sequences (Schweinfest et al, **Genet. Anal. Techn. Appl. 7**: 64 (1990); Sive and St. John, **Nucl. Acids Res. 16**: 10937

(1988)). Clones from this subtraction are listed in the column "Hippocampus Subtracted" in Table 1.

The EST sequences from this Example 1 are identified as SEQ ID NOs 1-315.

TABLE 1. cDNA Library Composition Determined By Random Clone Sequencing

	cDNA Library							
	Hippocampus		Hippocampus Subtracted		Fetal Brain		Temporal Cortex	
EST Category	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Databases Match--Human								
Mitochondrial Genes	48	12.8	10	8.6	3	7.9	6	7.5
Repeats: Alu, Line-1, etc.	39	10.4	14	12.2	6	15.8	0	0
Ribosomal RNA	10	2.7	7	6.0	0	0	11	13.8
Other Nuclear Genes	32	8.6	7	6.0	4	10.5	0	0
Database Match--Other	32	8.6	7	6.0	5	13.2	4	5.0
No Database Match	160	42.8	44	37.9	20	52.6	6	7.5
poly A Insert	53	14.1	24	20.7	0	0	27	33.7
No Insert	1	0.3	3	2.6	0	0	26	32.5

EXAMPLE 2

Sequencing of Additional ESTs: Second set

Over 2600 additional cDNA clones have been isolated, partially sequenced and screened. The clones were isolated from four human brain cDNA libraries. The new sequences thus discovered, together with the 315 brain ESTs from Example 1, correspond to over 2400 new human genes. These data represent an approximate doubling of the number of human genes identified by DNA sequencing.

Specifically, four cDNA libraries were used as sources of clones for sequencing. Human hippocampus and fetal brain libraries, plasmid template preparation, sequencing reactions, and automated sequencing were performed as described (Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B., Moreno, R.F., Kerlavage, A.R., McCombie, W.R., & Venter, J.C. **Science, 252**: 1651-56 (1991)). A pooled probe consisting of inserts from 10 different EST clones with sequences that matched either mitochondrial genes or the 18S or 28S ribosomal RNAs was used to prescreen a gridded filter array of the hippocampus library; nonhybridizing clones are referred to as the "prescreened library". Another fetal brain library was constructed by and was a gift from Bento Soares (Columbia University). A directionally-cloned library was prepared using the method of Rubenstein, et al. (Rubenstein, J., Elizabeth, A., Brice, A., Ciaranello, R., Denney, D., Porteus, M. & Usdin, T. **Nucl. Acids Res. 18**: 4833-4842) using human adult brain mRNA purchased from Clontech (Palo Alto, CA; Catalogue # 6516-1). Of 482 clones analyzed by restriction enzyme digestion, 33% contained inserts at least 1500 base pairs in length. Stratagene hippocampus and fetal brain library totals include data from Adams et al **Science 252**: 1651.

Sequences of nuclear-encoded cDNAs that did not include interspersed repeats (Schmid, C. W. & Jelinek, W. R. **Science**

216: 1065-1070 (1982); Paulson, K. E., Deka, N., Schmid, C. W., Misra, R., Schlinder, C. W., Rush, M. G., Kadyk, L., & Leinwand, L. **Nature 316**: 359-361 (1985); Fanning, T. G. & Singer, M. F. **Biochem. Biophys. Acta 910**: 203-212 (1987)) were searched against all of GenBank and, in 6-frame translation, against a comprehensive, non-redundant peptide database using the network BLAST (Altschul, S. F., Gish, W., Miller, W., Myers, E.W., & Lipman, D. **J. Mol. Biol. 215**: 403-410 (1990)) server at the National Center for Biotechnology Information. BLAST output was parsed, and an interactive alignment editor was used to select which matches, if any, from each search to record in a relational EST database, which was developed to track sequencing, identification, tissue localization, physical mapping, and the public distribution of the clones, mapping and sequence data. For significant similarities, a putative gene name and Protein Identification Resource (PIR) gene family identification (Barker, W., George, D., Hunt, L., & Garavelli, J. **Nucl. Acids Res. 19 (Suppl)**: 2231-2236 (1991)) for the EST were assigned. ESTs without significant matches using BLAST were searched in translation against PIR using FASTA. Ten additional marginal matches were found. A total of 2300 new EST sequences comprising 765,505 nucleotides from the current data set have been submitted to GenBank and assigned accession numbers M77851-M79278 and M85308-M86179. All ESTs except those multiply representing actin, tubulin, and myelin basic protein clones were submitted. ATCC accession numbers of cDNA clones from which ESTs were derived are 77501-78999 and 81000-81756. The Genome Data Base expressed D-segment numbers for these clones are D0S1E - D0S2422E. The ESTs from this Example are identified herein as SEQ ID NOs 316-2407.

EXAMPLE 3

EST Characterization: First Set

ESTs including SEQ ID NOs 1-315 were analyzed as follows. Initially, the EST sequences were examined for similarities in the GenBank nucleic acid database (GenBank Release 65.0), Protein Information Resource Release 26.0 (PIR), and ProSite (MacPattern from the EMBL data library, Fuchs R. **Comput. Appl. Biosci.** 7: 105 (1990) Release 5.0 were used). BLAST was used to search Genbank and the PIR (both maintained by the National Center for Biotechnology Information) ESTs without exact GenBank matches were translated in all six reading frames and each translation was compared with the protein sequence database PIR and the ProSite protein motif database. Comparisons with the ProSite motif database were done by means of the program MacPattern from the EMBL Data Library. GenBank and PIR searches were conducted with the "basic local alignment search tool" programs for nucleotide (BLASTN) and peptide (BLASTX) comparisons (Altschul et al, **J. Mol. Biol. 215**: 403 (1990)). PIR searches were run on the National Center for Biotechnology Information BLAST network service. The BLAST programs contain a very rapid database-searching algorithm that searches for local areas of similarity between two sequences and then extends the alignments on the basis of defined match and mismatch criteria. The algorithm does not consider the potential gaps to improve the alignment, thus sacrificing some sensitivity for a 6-80 fold increase in speed over other database-searching programs such as FASTA (Peqarson and Lipman, **Proc. Natl. Acad. Sci. USA, 85**: 2444 (1988)).

Sequence similarities identified by the BLAST programs were considered statistically significant with a Poisson P-value than 0.01. The Poisson P-value less than the probability of as high a score occurring by chance given the number of residues in the query sequence and the database.

After the BLASTN search, 30 unmatched ESTs were compared against GenBank by FASTA to determine if significant matches were missed due to the use of BLASTN for the database search. No additional statistically significant matches were found. Statistical significance does not necessarily mean functional similarity; some of the reported matches may indicate the presence of a conserved domain or motif or simply a common protein structure pattern. Those ESTs identified as fully corresponding to known human genes or proteins are not included in this disclosure. Statistically significant matches are reported in Table 2, together with the length and percent identity or similarity of each alignment.

On the basis of database searches, 609 EST sequences were classified into eight groups as shown in Table 1 (see Example 1 above). Four groups, with 197 or 32% of the sequences, consist of matches to human sequences: repetitive elements, mitochondrial genes, ribosomal RNA genes, and other nuclear genes. Forty-eight (8%) of the sequences matched non-human entries in GenBank or PIR while 230 (38%) had no significant matches. The remaining 134 (22%) sequences contained no insert or consisted entirely of polyA between the EcoRI cloning sites.

Thirty-six ESTs matched previously sequenced human nuclear genes with more than 97% identity. Four of these ESTs are from genes encoding enzymes involved in maintaining metabolic energy, including ADP/ATP translocase, aldolase C, hexokinase, and phosphoglycerate kinase. Human homologs of genes for the bovine mitochondrial ATP synthase F_0ß-subunit and porcine aconitase were also found (Table 2). Brain-specific cDNAs included synaptophysin, glial fibrillary acidic protein (GFAP), and neurofilament light chain. At least six ESTs are from genes encoding proteins involved in signal transduction: 2',3'-cyclic nucleotide 3'-phosphodiesterase (2 ESTs), calmodulin, c-erbA-α-2, $G_s\alpha$, and Na^+/K^+ ATPase α-subunit. Other ESTs were matches to genes for ubiquitous structural proteins -- actins, tubulins, and

fodrin (non-erythroid spectrin). ESTs also document the presence in the hippocampus cDNA library of the ret proto-oncogene, the ras-related gene rhoB, and one of the chromosome 22 breakpoint cluster region transcripts. Eight ESTs are from genes known to be associated with genetic disorders (Online Mendelian Inheritance in Man). More than half of the human-matched ESTs from Example 1 have been mapped to chromosomes, indicating the bias of GenBank entries toward well-studied genes and proteins.

ESTs without significant GenBank matches were also compared to the ProSite database of recognized protein motifs. Not counting post-translational-modification signatures, fifty-four sequences contained motifs from the database. Some patterns, particularly the "leucine zipper", are found in scores or hundreds of proteins that do not share the functional property implied by the presence of the motif.

Similarities to sequences from other organisms were also detected in the BLAST searches of GenBank and PIR (Table 2). Several ESTs displayed similarity to "housekeeping" genes, including the ribosomal proteins S10 and L30 (rat) and the above glycolytic enzymes. EST00257 (SEQ ID NO:77) shows strong nucleotide sequence similarity to the squid (67%) and Drosophila (70.4%) kinesin heavy chain. Kinesin was first described as a microtubule-associated motor protein involved in organelle transport in the squid giant axon (Vale et al, **Cell 42**: 39 (1985)). Six oncogene-related sequences were also among the cDNA clones sequenced. EST00299 (SEQ ID NO:180) and EST00283 (SEQ ID NO:271) show similarity to several ras-related genes and EST00248 (SEQ ID NO:102) matched the 3' untranslated region of the bovine substrate of botulinum toxin ADP-ribosyltransferase. Similarities with an S. cerevisiae RNA polymerase subunit and Torpedo electromotor neuron-associated protein were also observed. Two ESTs may represent new members of known human gene families: EST00270 matched the three ß-tubulin genes with 88-91% identity and

EST00271 (SEQ ID NO:248) matched α-actinin with 85% identity at the nucleotide level.

Among the most interesting of the primary sequence relationships was the similarity of ESTs to the Drosophila genes Notch and Enhancer of split. Nucleotide and peptide alignments of EST00256 (SEQ ID NO:188) and EST00259 (SEQ ID NO:227) with the Drosophila genes have been demonstrated. Both genes are part of a signal cascade encoded by the "neurogenic" genes that are involved in the differentiation of neuronal and epidermal cell lineages in the neuroectoderm of the developing Drosophila embryo (Campos-Ortega, **Trends in Neuro. Sci. 11**: 400 (1988)). It has been proposed that the Enhancer of split protein interacts with a membrane protein that is the product of the Notch gene to convert a developmental signal into an altered pattern of gene expression (id. **J. Mol. Biol. 215**: 403 (1990)). EST00256 (SEQ ID NO:188) matches near the 5' end of the Enhancer of split coding sequence, away from the mammalian G protein ß subunit- and yeast cdc4-like elements (Hartley et al, **Cell 55**: 785 (1988); Klambt et al. **EMBO J. 8**: 203 (1989)). Part of the EST00259 (SEQ ID NO:227) match to Notch in the cdc10/SW16 region that is similar to three cell-cycle control genes in yeast and is tightly conserved in the Xenopus Notch homolog, Xotch. In Drosophila, Enhancer of split is absolutely required for formation of epidermal tissue. Notch contains several epidermal growth factor-like repeats and appears to play a general role in cell-cell communication during development (Banerjee and Zipursky, **Neuron 4**:177 (1990)).

Seven genes were represented by more than one EST. Comparisons of all the ESTs against one another revealed two overlaps of unknown ESTs: EST00233 (SEQ ID NO:32) and EST00234 (SEQ ID NO:8) match in opposite orientations and EST00235 (SEQ ID NO:204) and EST00236 (SEQ ID NO:148) match in the same orientation beginning at the same nucleotide. Five human genes were represented by more than one EST: ß-

actin (3), λ-actin (2), α-tubulin (2), α-2-macroglobulin (2), and 2'3'-cyclic-nucleotide-3'-phosphodiesterase (2). Those few instances where two or more ESTs represent different portions of a single cDNA can be readily ascertained when the sequence of the full cDNA insert is determined in accordance with Example 13.

Example 4

EST Sequences Characterization: Second Set

The ESTs of Example 2, including SEQ ID NOs 316-2407, were screened against known sequences listed in GenBank and other databases, as in Example 3. The results are reported in Table 2. The quality of the match is given as percent identity and length in base pairs for nucleotide matches and amino acid residues for peptide matches. In many cases ESTs match multiple domains on several related proteins; for example, EST00825 matches two transmembrane domains on both GABA and Norepinephrine transporters. Nucleotide databases are: GenBank (GB), and EMBL (E); peptide databases are: GenPept (GPU), Swiss-Prot (SP), and PIR.

The great majority (83%) of the partial cDNA sequences reported in Example 2 are unrelated to any sequences previously described in the literature. Based on database matches to known genes from humans as well as from such evolutionarily distant organisms as *E. coli*, yeast, *C. elegans*, *Drosophila*, barley, *Arabidopsis*, rice, and green algae, we have preliminarily identified the functional type of a number of the ESTs (Table 2). These include a novel gene similar to Notch/Tan-1 (Adams et al., supra), a new neurotransmitter transporter gene, and a new member of the multi-drug resistance gene family. Several genes involved in development or cell differentiation in *Drosophila* are represented by similar human ESTs, including seven *in absentia* (Carthew, R. & Rubin, G. Cell 63: 561-577 (1990)),

big-brain (bib) (Rao, Y., Jan, L., & Jan, Y. **Nature 345**: 163-167 (1990)), the discs tumor suppressor (Woods, D. & Bryant, P. **Cell 66**: 1-20 (1991)), and the homeotic gene orthodenticle (Finkelstein, R., Smouse, D. Capaci, T., Spradling, A. & Perrimon, N. **Genes. Dev. 4**: 1516-1527 (1990)). New members of gene families previously known in humans include a Ca^{+2}-transporting ATPase, an ADP ribosylation factor, and a new neural-cell adhesion molecule gene.

The 1971 ESTs without a putative identification were analyzed using the coding-region prediction program CRM via the GRAIL server (Uberbacher, E. & Mural, R. **Proc. Natl. Acad. Sci. USA 88**: 11261-5 (1991)). Fifteen percent of the unknown ESTs scored an excellent probability of containing protein-coding sequence. Fifty percent of the ESTs to known human genes contain protein-coding sequences, therefore, at most half of the unknown ESTs are likely to contain coding sequences. We have found no evidence that genomic DNA or cDNA to unspliced precursor RNA is a major contaminant of either the hippocampus or fetal brain library.

Table 2: ESTs Identified by Database Matches

SEQ ID	EST#	Putative Identification	Accession	DB	Len	%ID
208	EST00250	60K filarial antigen	A28209	PIR	108	56.9
2320	EST01784	60K filarial antigen	A28209	PIR	88	50.6
969	EST01982	ADP-ribosylation factor 1	B33283	PIR	84	41.2
1834	EST01620	AMP deaminase, brain	A37056	PIR	57	100.0
97	EST00289	Aconitase	A35544	PIR	105	90.6
251	EST00370	Actin, other	S10021	PIR	44	51.1
248	EST00271	Actinin, alpha	HUMACTAR	GB	271	85.3
891	EST01891	Actinin, alpha	HUMACTAR	GB	315	81.6
1500	EST02538	Actinin, alpha	HUMACTAR	GB	271	75.0
132	EST00110	Agrin	RATAGR	GB	269	82.2
1852	EST01625	Agrin	RATAGR	GB	103	84.6
1094	EST02113	Ala	HUMALA	GB	92	82.8
691	EST00675	Alcohol dehydrogenase	RICGOS2G_1	GPU	38	59.0
2408	EST00244	Amyloid A4	HUMAFPA4	GB	135	91.9
1965	EST01664	Amyloid A4	A29030	PIR	52	54.7
2068	EST01694	Amyloid A4	QRHUA4	PIR	83	69.0
2092	EST01700	Anion exchanger homolog AE3	A33638	PIR	95	97.9
1880	EST01634	Axonal glycoprotein TAG-1	A34695	PIR	69	87.1
1492	EST02530	B cell-specific Mo-MLV integration site 1 (bmi-1)	MUSBMI1A	GB	111	87.5
1277	EST02306	Bib protein	S09699	PIR	57	53.4
13	EST00255	Cadherins	CADN$HUMAN	SP	41	45.2
1348	EST02378	cAMP-dependent protein kinase inhibitor	MUSPKI	GB	234	91.5
1931	EST01041	cAMP-regulated phosphoprotein	B35308	PIR	21	86.4
1413	EST02447	cAMP-specific phosphodiesterase	HUMPDEAA	GB	363	69.0
396	EST01443	CDPdiacylglycerol-serine O-phosphatidyltransferase	JH0368	PIR	33	41.2
1956	EST01663	Ca2+-transporting ATPase 2	B28065	PIR	125	88.9
1126	EST02146	Calbindin D28	RATCALBD28	GB	81	87.8
1039	EST02055	Calcium channel	S05054	PIR	33	67.6
1910	EST01645	Calmodulin	RATRCM1	GB	120	90.1
485	EST01466	Calmodulin-dependent protein kinase, type II, beta	A26464	PIR	93	98.9
913	EST01913	Clathrin coat assembly protein AP50 homolog	YSCYAP54_1	GPU	62	63.5
2004	EST01676	Cofilin	PIGCOFIL	GB	132	89.5
2400	EST01824	Cysteine-rich intestinal protein	GYRTI	PIR	56	66.7
1588	EST02633	D22Z3 repetitive DNA	HUMREP	GB	160	76.4
2192	EST01257	Diacylglycerol kinase, lymphyocyte	S09156	PIR	44	42.2
1441	EST02477	Diamine acetyltransferase	ATDA$HUMAN	SP	74	45.3
650	EST00642	Dilute (myosin heavy chain)	MUSDILUTE_1	GPU	27	100.0
2302	EST01779	Discs-large tumor suppressor	DRODLGA_1	GPU	53	63.0
188	EST00256	Enhancer of split	A30047	PIR	86	58.6
2289	EST01325	Fatty acid synthase	RATFAS	GB	98	79.8
310	EST00377	Fo ATPase beta subunit, mitochondrial	BOVMTASB	GB	293	85.4
1332	EST02362	GA binding protein, beta subunit	MUSGAC_1	GPU	86	90.8
1667	EST00825	Gamma-aminobutyric acid transporter	A35918	PIR	26	59.3
2217	EST01738	Gelation factor ABP-280	A37098	PIR	74	80.0
1412	EST02446	Glutamate-aspartate carrier protein	JV0092	PIR	57	37.9
1020	EST02034	Glutaminase	GLS$RAT	SP	34	74.3
1885	EST01639	Histocompatibility antigen modifier 1	A37779	PIR	63	75.0
1495	EST02533	Hypothetical 43.5K protein	JU0319	PIR	43	52.3
2326	EST01791	Inositol-1,4,5-trisphosphate 3-kinase	JN0129	PIR	65	68.2
SEQ ID	EST#	Putative Identification	Accession	DB	Len	%ID

724	EST01529	Interferon-induced 54K protein	IN54$HUMAN	SP	76	70.1
1035	EST02051	J1 protein	MUSJ1PRO	GB	382	85.7
1229	EST02258	KUP protein	HUMKUPMR_1	GPU	54	36.4
993	EST02007	Kinase 5 protein	CHKCEK5_1	GPU	68	94.2
77	EST00257	Kinesin	A35075	PIR	57	86.2
78	EST00258	Kinesin	A35075	PIR	62	47.6
2245	EST01748	Kinesin	A35075	PIR	98	52.5
2282	EST01764	Lamin B receptor	A38427	PIR	76	71.4
2173	EST01724	Lon protease	JQ0901	PIR	103	41.3
1427	EST02463	Long-chain-fatty-acid-CoA ligase	A36275	PIR	36	62.2
313	EST00276	Lysosomal membrane glycoprotein 1 (LAMP-1)	A31959	PIR	53	46.3
161	EST00247	MARCKS (myristoylated alanine-rich protein kinase	BOVMARCKS	GB	139	83.6
1386	EST02418	MARCKS homolog	MMF52	EU	237	92.4
769	EST00734	MARCKS homolog	S08341	PIR	61	40.3
43	EST00371	Maternal G10 protein	S05955	PIR	38	92.3
1468	EST02505	Matrin 3	RATMATRIN3	GB	137	93.5
639	EST00832	Membrane transport superfamily (GTP-dependent)	A24400	PIR	63	39.1
1894	EST01643	Membrane transport superfamily (GTP-dependent)	A24400	PIR	71	50.0
824	EST01865	Microtubule-associated protein 1B	RATNEU	GB	293	86.4
223	EST00368	Microtubule-associated protein 1B	A33845	PIR	30	54.8
2032	EST01683	Microtubule-associated protein 1B	A33845	PIR	49	62.0
2017	EST01678	Milk fat globule membrane protein	A36479	PIR	48	61.2
1704	EST01580	Myeloid differentiation primary response gene MyD1	MUSMYD118_1	GPU	76	88.3
2226	EST01744	NAD(P)+ transhydrogenase (B-specific)	DEBOXM	PIR	88	93.1
1567	EST02610	Neural cell adhesion molecule L1	S05479	PIR	82	43.4
506	EST01471	Neurexin	S06017	PIR	120	84.3
1566	EST02609	Neutrophil oxidase factor	A34855	PIR	43	47.7
952	EST01961	Notch/Xotch	HUMTAN1_1	GPU	85	57.0
227	EST00259	Notch/Xotch	A35844	PIR	74	85.3
1395	EST02429	Nuclear factor 1-like protein (NF1)	HAMNF1A	GB	111	92.0
1681	EST01573	Nucleoside diphosphate kinase	A33386	PIR	71	52.8
346	EST01828	Otd homeotic protein	A35912	PIR	35	52.8
2254	EST01751	Phosphatidylinositol-4,5-bisphosphate phosphodiest	A28807	PIR	40	90.2
1869	EST00992	Polymyxin B resistance	A32714	PIR	20	76.2
93	EST00287	Processing enhancing protein	S03968	PIR	96	58.8
2353	EST01806	Prohibitin	RATPROHIB_1	GPU	120	97.5
2297	EST01775	Prohormone cleavage enzyme	MUSMPC1A_1	GPU	91	93.5
9	EST00376	Prolyl endopeptidase	PIGPREP	GB	223	83.9
1069	EST02087	Protein kinase C, zeta	HUMPKCL	GB	382	58.7
1933	EST01650	Protein phosphatase 2A beta subunit	HUMPROP2AB	GB	288	76.8
202	EST00298	Protein-tyrosine phosphatase LRP	LRP$MOUSE	SP	62	44.4
1654	EST01572	Protochlorophyllide reductase	S04783	PIR	34	57.1
38	EST00374	RNA polymerase II 6th subunit (RPO26)	A36352	PIR	72	75.3
1478	EST02515	Rab5	F34323	PIR	91	82.6
2368	EST01389	Radial spoke protein 3	S05962	PIR	58	52.5
37	EST00038	ras p21-like small GTP-binding protein (smg GDS)	BOVSMGGDS	GB	131	89.4
180	EST00299	ras-related proteins	S10493	PIR	51	46.1
1700	EST01579	Retrovirus-related gag polyprotein	FOHUE2	PIR	95	77.1
1511	EST02550	Retrovirus-related pol polyprotein	GNLJGL	PIR	50	54.9
102	EST00248	rho H12/ ARH12	BOVBGBRH	GB	195	79.6
1715	EST01583	Ribosomal protein L18a	R5RT18	PIR	68	95.7

SEQ ID	EST#	Putative Identification	Accession	DB	Len	%ID

1856	EST01627	Ribosomal protein L1a	A24579	PIR	75	63.1
1974	EST01667	Ribosomal protein L3	JQ0771	PIR	74	80.0
301	EST00300	Ribosomal protein L30	R6RT30	PIR	57	96.5
22	EST00301	Ribosomal protein S10	R3RT10	PIR	66	97.0
2402	EST01826	Ribosomal protein S10	R3YM10	PIR	36	51.4
463	EST01459	Ribosomal protein YL10	S11581	PIR	40	68.3
1408	EST02442	Seven in absentia	A36195	PIR	46	80.8
299	EST00249	smg p25A GDP dissociation inhibitor	A35652	PIR	97	77.5
951	EST01960	Spectrin, beta	HUMSPTB	GB	268	67.7
2089	EST01699	Sperm membrane protein	A35981	PIR	52	58.5
2073	EST01697	Succinate dehydrogenase flavoprotein	BOVSDHFP1_1	GPU	44	100.0
2138	EST01715	Succinate dehydrogenase flavoprotein	BOVSDHFP1_1	GPU	49	92.0
430	EST00472	Synaptotagmin (p65)	SY65$HUMAN	SP	27	53.6
1371	EST02402	Talin	MUSTALINR_1	GPU	79	81.2
1771	EST01601	Thiosulfate sulfurtransferase (rhodanese)	ROBO	PIR	65	81.8
300	EST00232	Transforming protein (dbl)	TVHUDB	PIR	25	65.4
189	EST00282	trkB	A35104	PIR	33	67.6
653	EST01512	Tubulin, alpha	HUMTUBAG	GB	223	75.0
594	EST01490	Tubulin, beta	HUMTBB5	GB	298	93.6
757	EST01542	Tubulin, beta	HUMTUBBM	GB	217	90.4
1245	EST02274	Tubulin, beta	A26561	PIR	105	88.7
1147	EST02169	Tyrosine kinase	HUMECK	GB	384	74.3
1701	EST00853	Unc-104	JN0114	NR	36	45.0
2121	EST01711	Valine-tRNA ligase	A29871	PIR	56	57.9
187	EST00152	Wilm's tumor-related protein	HUMOM	GB	228	99.6
1726	EST01588	XPR2 alkaline extracellular protease	B26955	PIR	88	46.1
249	EST00275	Zinc Finger Proteins	S06551	PIR	25	57.7
413	EST01446	Zinc Finger Proteins	S00754	PIR	45	60.9
469	EST01460	Zinc Finger Proteins	C32891	PIR	34	54.3
833	EST01560	Zinc Finger Proteins	S00754	PIR	105	67.0
1230	EST02259	Zinc finger proteins	S00754	PIR	71	62.5
1496	EST02534	Zinc finger proteins	A34612	PIR	50	45.1
2324	EST01352	Zinc Finger Proteins	S10397	PIR	29	56.7

There is little redundancy in EST sequencing according to the present invention. Of the nuclear-encoded messenger RNAs, the most common ESTs were to the β-actin (0.6% of the EST clones) and myelin basic protein genes (MBP, 0.5% of the clones). MBP, a highly expressed structural component of nerve tissue (Kamholtz, J., de Ferra, F., Puckett, C., & Lazzarini, R. **Proc. Natl. Acad. Sci., USA 83**: 4962-4966 (1986)), displays four alternate splicing forms, of which at least two are present among the ESTs reported here. Other common ESTs were Gs-alpha gamma-actin and both a- and alpha-tubulin.

By matching ESTs to known database sequences, a phenotypic characterization of the tissue begins to emerge. Protein superfamilies matched by ESTs were grouped into three broad functional categories to assess the biological spectrum represented by these randomly selected cDNA clones. Structural and metabolic classes comprised about 30% of the ESTs with database matches. Twenty-five percent were involved in regulatory pathways and the remainder were not classifiable. Eleven of the eighteen enzymes of glycolysis and the citric acid cycle are represented by at least one subunit or isozyme. In addition, several genes not previously known to be expressed in the brain were matched, including spermine/spermidine acetyltransferase (Casero, R., Celano, P, Ervin, S., Applegren, N., Wiest, L. & Pegg, A. **J. Biol. Chem. 266**: 810-814 (1991)) and osteopontin (Young, M., Kerr, J., Termine, J., Wewer, U., Wang, M., McBride, W. & Fisher, L. **Genomics 7**:491-502 (1990)).

EXAMPLE 5

Mapping of ESTs to Human Chromosomes

Randomly selected ESTs corresponding to SEQ ID NOs. were assigned to chromosomes via PCR (see Table 3). Oligonucleotide primer pairs were designed from EST

sequences to minimize the chance of amplifying through an intron. The oligonucleotides were 18-23 bp in length and designed for PCR amplification using the computer program INTRON (National Institutes of Mental Health, Bethesda, MD). The program is based on the assumptions that: 1) introns are genomic sequences that interrupt the coding and noncoding sequences of genes (Smith, **J. Mol. Evol. 27**:45-55 (1988)); 2) there are consensus sequences for splice junctions (Shapiro, et al., **Nucl. Acids Res. 15**:7155-7174 (1987)); and 3) that 90% of the human genes studied have 3' untranslated regions of mRNA not interrupted by introns in the genomic DNA (Hawkins, **Nucl. Acids Res. 16**:9893-9908 (1988)).

The program evaluates the likelihood that a given GG or CC dinucleotide represents a former exon-intron boundary. Specifically, every input strand is processed by the INTRON program twice, first evaluating the sense mRNA strand, and then processing the complementary or anti-sense strand. The program evaluates each sequence by finding all GG or CC pairs (possible former splice sites), searching for STOP codons in all three reading frames, and analyzing the GG or CC pairs surrounded by stop codons. All regions of the EST that are unlikely to contain splice junctions based on CC content, GG content, and stop codon frequency are then marked by the program in uppercase.

The creation of PCR primers from known sequences is well known to those with skill in the art. For a review of PCR technology see Erlich, H.A., **PCR Technology; Principles and Applications for DNA Amplification**. 1992. W.H. Freeman and Co., New York. ESTs were examined for the presence of stop codons in each reading frame and for consensus splice junctions. The presence of stop codons and absence of splice junction sequences are more characteristic of 3' untranslated sequences than of introns. The untranslated sequences are unique to a given gene; thus, primers from

these regions are less likely to prime other members of a gene family or pseudogenes.

The primers were used in polymerase chain reactions (PCR) to amplify templates from total human genomic DNA. PCR conditions were as follows: 60 ng of genomic DNA was used as a template for PCR with 80 ng of each oligonucleotide primer, 0.6 unit of Tag polymerase, and 1 uCu of a ^{32}P-labeled deoxycytidine triphosphate. The PCR was performed in a microplate thermocycler (Techne) under the following conditions: 30 cycles of 94°C, 1.4 min; 55°C, 2 min; and 72°C, 2 min; with a final extension at 72°C for 10 min. The amplified products were analyzed on a 6% polyacrylamide sequencing gel and visualized by autoradiography. If the size of the resulting product was equivalent to the EST from which the primers are derived, then the PCR reaction was repeated with DNA templates from two panels of human-rodent somatic cell hybrids; BIOS PCRable DNA (BIOS Corporation) and NIGMS Human-Rodent Somatic Cell Hybrid Mapping Panel Number 1 (NIGMS, Camden, NJ).

PCR was used to screen a series of somatic cell hybrid cell lines containing defined sets of human chromosomes for the presence of a given EST. DNA was isolated from the somatic hybrids and used as starting templates for PCR reactions using the primer pairs from EST sequences selected above. Only those somatic cell hybrids with chromosomes containing the human gene corresponding to the EST will yield an amplified fragment. ESTs were assigned to a chromosome by analysis of the segregation pattern of PCR products from hybrid DNA templates. For a review of techniques and analysis of results from somatic cell gene mapping experiments. (See Ledbetter et al., **Genomics** 6:475-481 (1990).) The single human chromosome present in all cell hybrids that give rise to an amplified fragment represents the chromosome containing that EST.

The assignment of 100 ESTs and corresponding genes to chromosomes by PCR is shown in Table 3.

Table 3: Assignment of ESTs to Chromosomes by PCR

SEQ ID	EST#	Chr	PRIMER #1	PRIMER #2
5	EST00012	1	TCCAGGCAATCCCAGAATAG	CTAATTGAGCTCACTGGCCC
57	EST00058	1	CTGTTTGCAAGTTTCAAAGC	GCCATTTCTAACAACCAGAG
64	EST00066	1	GCCATTGTGCTGAATAGAGT	GTTAGTGTTTCCTTAGCAAG
83	EST00079	1	CAGCTAATTGACCTGGGCTA	CAACATGCTCTGAGCTTTAG
83	EST00079	1	GGCAGAGCATAATGAGTATA	CATATGCATATGGTCCCTAT
91	EST00086	1	AGTTTAGATGGAGGGCTGTC	TCTGCCCTAATGCGCAGGCT
105	EST00365	1	CTTAATCACCTCCCTTTTGT	CCTTAGTTGGAGATAAGGTC
109	EST00095	1	AGTCTAATCCTGTACACTTG	CGGGCTTTCTCTGAATTGGT
116	EST00100	1	TTAGAAGTGCCCATGGGAGG	TTTTAAGGCTCTGGAGTGTT
141	EST00118	1	CTCAGAGAAACTTAGGTGAA	CTACAGAATCATTTCACCAG
220	EST00372	1	AAGTTGCACATTGCCCAAGG	ATAGTACTGCAAGGTTATTC
237	EST00187	1	TTACAAATTTCTCTTGACGC	CTGAAGGAGCACAGTTTCTC
242	EST00192	1	GGATCAGATAATCAAACAGG	GCTTAGGATATGAATGCATA
259	EST00202	1	GCATCACAGTTTAACTGAGG	CTACATATTTGTGCCTCCTT
269	EST00293	1	CTGTTGCTGTGCAGTAGCTT	CTTTTGACCCAGTGAAACTT
299	EST00249	1	GATCATGCAGACGTAGATAT	CCAACTCCTGCCAGATCATT
1651	EST00810	1	TAGTCGCTGTAAGTTGATTC	GCTTTGCTGGATGCTTCATT
16	EST00021	2	CAGGCAAGTTTCTTCCAGGA	TCAGACCCATGGTCAGCTT
1898	EST01013	2	GGCTGAGAACGGTTAGCATA	CCCTCAGCTTAGGGGAATG
8	EST00234	2	TAGAAGGCAAACTATGTCCC	GGTTGAGGATTGGCTTTTAC
36	EST00037	2	AGCCAGAAGGCTGCTTAAAG	GCAGTGAACCAGTACTCCTA
123	EST00106	2	GTCTAATTTGTAACCTTCAG	GATAGATTGTATAAGAAGCC
192	EST00155	2	GATTTATGTCTGGGAACTAA	GCAGCATGTGAAAGAATGAT
200	EST00162	2	TTTAATGGGTGGTGGGAGCT	CGATGCACATCCTTCTCCAT
284	EST00216	2	CCTAAGAATTCGTTTGGCTC	GTCTGGCACATAATAGATTTG
102	EST00248	3	ATACTACATCTAGTCTGG	TTACAGTTCTGTGGTTTC
167	EST00138	3	AAACAGCTGCGGAGTACA	AAAGGATCCTCCACTCCAGA
12	EST00274	3	CCTAGCAAACTCATACACAC	CATAAGTGAATGGACACAGG
60	EST00062	3	ACACATTAACGGTGCTGCAG	GGAATCAGCCCTTGAGGACT
77	EST00257	3	AAGCTCACAACGCAGATCTG	CTGGAACAGCTTACAAAGGT
107	EST00093	3	ATTGAACTCTGTCAACAGTG	TGTAAAACAAAGGCCAAACT
108	EST00094	3	AL2-GCAGGATGTCAGTCTTTTGAG	AGCACACATTATCTACCACGGC
1706	EST00857	3	AL2-GCAGGATGTCAGTCTTTTGAG	CCAGCACACATTATCTACCACG
37	EST00038	4	AACTTCGCAGTCATGAGAAC	TGTATCGGGCAGTTCTCAG
6	EST00013	4	CACATGTTCTCCCTCTTTCA	GCATTTTGGAGCTCTTCCGT
37	EST00038	4	AL2-GGAAGTACAGGATTTGGC	TTAGAGATGGGATGATGCCG
31	EST00033	5	TGGGTACCCTAAGGTGTTTG	GACTAATCTAAGGTCTAGG
28	EST00030	5	AGATAAGTTAGGAAGCTGGT	ACTCACTGCTAGTATCATCC
59	EST00061	5	AAAGTTTCTTAGCACCCCCC	CAGACTTTGACAAAAGAATC
74	EST00073	5	ATCAGACACGTGGCAGGGTT	AAGTCCCTGAGGGTGCAGAA
121	EST00104	5	TGAAGGCAGCTGCTAAATCT	GGATGTATTGATCTGACTCA
149	EST00123	5	ATACTGTCAACGGAGGGTGA	GTCTGCAGGTTTCTCCTTGA
235	EST00185	5	TTACTGTCCCATCAGATATC	TACACTCTTAAGAAGGTATG
1643	EST00803	5	GAGCGTTTAAAAGAGATTCT	TACAGACAGCCATGTTCCAA
1677	EST00835	5	AL2-TCTCCAACACAGTCATGC	CGGATGCCATCATATACC
23	EST00026	5	CCTGCAGTGACACTTAACAT	CTGCTCACCTGAAATTGATAC
121	EST00104	5	AL2-CAGATCAATACATCCTCTGGG	CTGTGCAGTGGTGAGTAAAAGG

SEQ ID	EST#	Chr	PRIMER #1	PRIMER #2
1	EST00007	6	TAGTTGATGGTCTGGGTTAT	GAAATCCCAGGGAGACAATG
19	EST00023	6	CAACTTACATTAGGGGTTTG	GACCTCATTAGAAGAGCCCA
155	EST00129	6	GGAAGCTGCCATATAAGCTC	TCAGTGTCGTACAATCTACC
224	EST00356	6	GCTGTATGTTAACCCTTTGT	TGGAACCCTCAAACACTGCT
288	EST00219	6	ACTTTCATGTTGAGAAGTAT	ATCTAGCTGAAACATTGCTG
1638	EST00798	6	CTTCATCTGTTAACTGTTGA	TGAAAATGAGTCACAGGCAG
1675	EST00833	6	AL2-ACCCAGTTCTCAAAGACC	GGTTTACCATTCAGAGGC
22	EST00301	6	CTCCGTGATTACCTTCATCT	TTGTAGGTATCTCTGTCAGCT
207	EST00167	7	GGTGCTACTTTGTGAATGCT	AGCAATGTGATTTTGTAGG
137	EST00272	7	AGTGGTCACTATCTACATGG	GATTCAGAATTACTAAGCCG
1659	EST00817	7	TGTATAGGCTCTACATAAAG	CTTAATCATGGATTCTTCGT
1680	EST00838	7	AL2-GTTCTTTCCCAGGTATGC	TTGTTGGTACTGAGGAAGTGCG
292	EST00223	8	TGCAGCAGTGACCATGAGAA	ATCATCTTTCCACGCGGCTT
134	EST00375	9	TCTGGGCTTCTGTGGTTCAA	CTGGCTGCTCAGCAACTCAT
1906	EST01021	9	GGATGTTTTCTATGTGACGA	TTCCAGTGCCCCTTTTGTCC
1645	EST00804	10	CTCCTTTGGGACAAACAACT	CCAACCCAAACATATTCTA
20	EST00024	10	AGCTGTTCCTGAGAGATGCA	CCTTGTGAAGAAAGACTTTC
157	EST00131	10	TCAGCAACAGGTCACTTTGG	CTAAGCATCTGCATGTCCAG
172	EST00142	10	TACTAGCATTTCTTACTCTC	TATGCTGATTGTTTGCACTC
250	EST00197	10	GGTGATTAGAGAGTCTGTTG	GAACTCTGTAGTGTTCTAAA
133	EST00111	11	GGAAATTAGGCTTAGCTCAC	GTGCAGAATACTTAGAGTCC
178	EST00294	11	GTTTGAAGGAAGTGATTTCC	TAGGGCCACCTCCAGTTCAT
10	EST00016	11	GTCTTTGGATTCTACGTAGA	CGATAATGACATTTCTTCTGG
126	EST00109	11	AL2-CTAACCACAACCCACACATTG	CCTCAGCACAAGAGAAGAATGG
7	EST00014	12	AACTTGCAACATAAATACTAG	GAGCAATGATTTCTAACAGT
254	EST00200	13	TTGTGTACTGTCTGATAGAC	TAAGCCATGGGCATCTATAA
2409	EST00273	13	GCAAGATGATGGAACATCCC	TTCCTTCTGGAGGCTCTACA
170	EST00295	14	GGTGCTTAAGGCCACTTTTG	CTTAGAGGATCATAGGTCTG
255	EST00201	14	CCAGGAGAGTAAGAAGATCA	GCAGAGTTGAATATGAACCT
290	EST00221	14	GTGCCAAGATGGCTCATGTA	GTATAGCTTTAAGCCAGTTC
293	EST00224	14	AATGCATTATGCCTGGTCTT	GGAAAAGTCTAGAACTTAGT
1664	EST00822	14	GGGTCAGAATTAAGAGGTCT	GTTCATCTCTAACTCCTTTC
315	EST00008	14	AAGCTGGCTGGGAAATGTTC	GTCATGCTAGTAAACTTACAC
1689	EST00845	14	AL2-AGGAGGAAGCTGAAATCC	GGAAGTCCATAAGAGACTCACC
95	EST00088	15	GTGACAGACCATGTCTATTG	AAGTGAGCGATTGCACCTTC
205	EST00165	15	AGGATGACCTGAGTGAGCTG	CCATGGCAGCAAGGAACTCT
33	EST00034	16	TGTGTGAAAGGGAGTCTTGT	CCATTTTGACTGTTCCATAG
247	EST00279	16	TGGCTAGGGCAGGCCTTAAA	GAGAAGAATATCAAATGGGG
18	EST00373	16	CCATCTGTGTCCCAATTAAGC	AGGGAAGAAGTCTAGAGCGA
68	EST00068	17	CAAAGACGGGAGACGAATGA	AGTGGAACGCGTGGCCTATG
1652	EST00811	17	GAGCTGCATGTTGATAAGTA	TTGACTTAAGCTGACCTTAA
1702	EST00854	17	AL2-TTGCTGTGGAATCCATGAGAG	GGCAAGTGATCTGTTCTTGG
84	EST00080	19	AGAGATGTCAGTCCATTATC	CTATTCCACCTTACTCAAGG
223	EST00368	19	CATCATGTCGGAGACGCATT	TGGATGACCTGAGTCTGCAG
21	EST00025	20	AGTTCTGGAGGCTAGGAGTT	ATGTAAGGACCCCTAGATGG
210	EST00168	20	TGTCAACTTCCCTTTGGCCT	GAAGCTTGCTCATTCAGGAA
136	EST00113	20	AL2-TCGGAGAAGTTGCAGTTTCTG	GTTAAAAGCTGTTAGACGGGGC
120	EST00103	22	CACTGACTGACTCCTCTTTA	GGAACCGTAACTCTCCATAG
313	EST00276	X	ATTGACCTTCAATGTAATAA	TTGGATTGGGCAAAATAG

SEQ ID	EST#	Chr	PRIMER #1	PRIMER #2
162	EST00133	X	ATGTGAGCATCTATACCTGC	AATGAAGGCATGAGAATAGG
1669	EST00827	X	CGGACAACTAGGATAAATGC	TACGCGTTTGAATGGCTTGA
1917	EST01029	X	GAATAGCATTATTAGCCAGT	GGACCTATTGGAGATCTACT
1708	EST00858	X	AL2-AAGGCGAGGATTATGTGC	TTCTACTGGGTACACTTCGACC

Abbreviation: AL2: Amino-Link-2 Fluorescent Tag, Chr.: Chromosome.

The foregoing techniques have been used to further localize 9 ESTs and their associated genes to precise locations onto chromosome 6 or chromosome X, as reflected in Table 4A (in Example 7 below), using sublocalization techniques that employ somatic cell hybrids. ESTs were used as hybridization probes and mapped to other chromosomes using techniques disclosed in Example 7. Somatic cell hybrids were prepared that contained defined subsets of chromosomes 6 and X. Methods for preparing and selecting somatic cell hybrids are known in the art. For a review of an exemplary procedure to generate somatic cell hybrids containing the short arm of human chromosome 6, see Zoghbi, et al., **Genomics 9(4):713-720 (1991)**. For a general review of somatic cell hybridization see Ledbetter et al.(supra). The hybrids were processed to obtain DNA and analyzed by PCR and by fluorescence in situ hybridization. SEQ ID NOs 19, 22, 1, 224, 288 mapped to chromosome 6, while SEQ ID NOs 162, 1917, 1699 and 1899 mapped to chromosome X using somatic cell hybrids.

EXAMPLE 6

Mapping of All ESTs to Human Chromosomes

The procedure of Example 5 is repeated for all of the ESTs from Examples 1 and 2 not previously mapped to human chromosomes. Data are generated corresponding to the data in Table 3 for all of the unmapped ESTs. As previously mentioned, virtually all of the ESTs will map to a unique chromosomal location. The inability of any ESTs to localize to a unique location will be readily ascertainable during the mapping process.

Physical mapping of the type reported in Table 4 on all the EST clones reported here would provide human chromosome markers spaced on average every 1.2 megabases and would roughly double the number of expressed sequences that have been localized to chromosomes (McKusick, V. **FASEB**

J. 5: 12-20 (1991)). Mapped ESTs are also a new resource to identify candidates for the estimated 5000 single-locus disease-associated genes (Id.).

EXAMPLE 7

Alternative Technique for Mapping to Chromosomes
Mapping of ESTs to chromosomes using fluorescence in situ hybridization

This technique was used to map an EST to a particular location on a given chromosome. Cell cultures, tissue, or whole blood were used to obtain chromosomes.

0.5 ml. of whole blood was added to RPMI 1640 and incubated 96 hours in a 5%CO_2/37°C incubator. 0.05 ug/ml colcemide was added to the culture one hour before harvest. Cells were collected and washed in PBS. The suspension was incubated with a hypotonic solution of KCl added dropwise to reach a final volume of 5 ml. The cells were spun down and fixed by resuspending the cells in methanol and glacial acetic acid (3:1). The cell suspension was dropped onto glass slides and dried.

The slides were treated with RNase A and washed then dehydrated in a series of increasing concentrations of ethanol.

The EST to be localized was nick-translated using fluorescently labeled nucleotide (Korenberg, Jr., et al., **Cell 53(3)**:391-400 (1988)). Following nick translation, unincorporated label was removed by spin dialysis through Sepharose. The probe was further extracted with phenol-chloroform to remove additional protein. The chromosomes were denatured in formamide using techniques known in the art and the denatured probe was added to the slides. Following hybridization, the cells were washed. The slides were studied under a fluorescent microscope. In addition, the chromosomes can be stained for G-banding or Q-banding using techniques known in the art.

The resulting metaphase chromosomes had fluorescent tags localized to those regions of the chromosome that were homologous to the EST. Thus, a particular EST was localized to a particular region on a given chromosome. In this manner, SEQ ID NOs 396, 485, 506, 1880 and 1894 were mapped using fluorescent in situ hybridization to locations on chromosomes 17, 7, 10 and 1 respectively (See Table 4B below). For a review of the technique see Verma et al., **Human Chromosomes: A Manual of Basic Techniques**. Pergamon Press, NY (1988), which is hereby incorporated by reference.

Table 4: Precise Chromosomal Localization of ESTs

	SEQ ID	EST#	Map Location
A.	19	EST00023	6p
	22	EST00301	6p
	1894	EST01643	6p21
	1	EST00007	6q
	224	EST00356	6q
	288	EST00219	6q
	162	EST00133	Xp11.21 - Xp21.2
	1917	EST01029	Xp11.21 - Xp21.2
	1669	EST00827	Xq26 - Xq27.1
	1899	EST01014	Xq28
B.	1880	EST01634	1q32
	485	EST01466	7p13
	506	EST01471	10q11.2
	396	EST01443	17q25

EXAMPLE 8

Automated DNA Sequencing Accuracy

ESTs that match human sequences in GenBank are excellent tools for the analysis of the accuracy of double-strand automated DNA sequencing. Ninety EST/GenBank matches were examined for the number of nucleotide mismatches and gaps required to achieve optimal alignment by the Genetics Computer Group (GCG) program BESTFIT (Devereux et al, **Nucleic Acids Research** 12: 387 (1984)).

The number of mismatches, insertions and deletions was counted for each hundred bases of the sequence (Table 5). As expected, the sequence quality was best closest to the primer and decreased rapidly after about 400 bases. The number of deletions and insertions relative to the GenBank reference sequence increased five- to ten-fold beyond 400 bases, while the number of mismatches doubled. The average accuracy rate for individual double-stranded sequencing runs was 97.7% to 400 bases.

TABLE 5. Accuracy Of Single-Run Double-Stranded Automated Sequencing

Bases from Primer	Mismatches/ Ambiguities[+]	Gaps Insertions[+]	Percent Deletions[+]	Aligned Accurate	Bases
101 - 200	1.45	0.18	0.19	98.2	8,800
201 - 300	1.72	0.25	0.11	97.9	8,130
301 - 400	2.07	0.98	0.37	96.6	5,404
>400	3.53	2.63	1.06	92.8	3,197

ESTs statistically identical to known human sequences and those matching mitochondrial and ribosomal genes were aligned with sequenced from GenBank using the GCG program BESTFIT. The first 85 nucleotides was polylinker sequence which was not aligned with the pBluescript SK reference sequence. Tabulation of errors began 15 bases into the BESTFIT alignment and thus is reported beginning with bases 101-200. [+]Error rates are reported as number of mismatches, insertions, or deletions per hundred aligned bases. "Mismatches" includes ambiguous base calls.

EXAMPLE 9

Probability of ESTs Containing Coding Sequences

The ESTs of the present invention were statistically evaluated using the coding-region prediction program CRM via the GRAIL server (Uberbacher, E. & Mural, R. **Proc. Natl. Acad. Sci. USA, 88**: 11261-5 (1991)). The CRM program uses a neural network to combine results from several different coding regions by looking at different 6 bp sequences found in coding exons and in introns. The program additionally conducts reading frame searches and assesses randomness at the third position of codons. This protocol categorizes sequences as having an excellent, good, marginal, or poor probability of containing coding regions. The results are reported in Tables 6-9. There were 219 ESTs categorized as "excellent" (Table 6); 120 categorized as "good" (Table 7); 113 categorized as "marginal" (Table 8); and 1743 categorized as "poor" (Table 9). These results indicate that most ESTs of the present invention comprise noncoding regions.

Table 6: ESTs with Excellent Probability of Containing Coding Sequence

SEQ ID#	EST#
7	EST00014
15	EST00020
48	EST00291
62	EST00064
66	EST00067
75	EST00074
98	EST00260
106	EST00092
108	EST00094
114	EST00098
115	EST00099
124	EST00107
128	EST00252
156	EST00130
164	EST00135
166	EST00137
174	EST00296
179	EST00145
183	EST00148
201	EST00163
205	EST00165
215	EST00172
230	EST00181
253	EST00199
263	EST00203
268	EST00369
270	EST00207
271	EST00283
273	EST00208
276	EST00211
281	EST00214
285	EST00286
333	EST00394
336	EST00397
339	EST00400
362	EST00418
389	EST00440
441	EST00481
454	EST00493
476	EST00509
493	EST00522
504	EST00529
516	EST00538
518	EST00540
551	EST01482
552	EST00565
559	EST00570
582	EST00592
602	EST00606
606	EST00609
608	EST00611
621	EST00620
635	EST00629
642	EST00634
644	EST00636
687	EST00671
700	EST00683
743	EST00714
753	EST00721
760	EST00726
764	EST00729
808	EST00761
823	EST01864
834	EST00771
886	EST01886
919	EST01921
930	EST01933

SEQ ID#	EST#
936	EST01939
948	EST01957
965	EST01978
973	EST01987
979	EST01993
980	EST01994
986	EST02000
1000	EST02014
1004	EST02018
1007	EST02021
1018	EST02032
1021	EST02035
1034	EST02050
1047	EST02063
1090	EST02109
1096	EST02115
1115	EST02135
1118	EST02138
1129	EST02149
1133	EST02153
1141	EST02163
1163	EST02187
1183	EST02208
1243	EST02272
1264	EST02293
1265	EST02294
1266	EST02295
1287	EST02317
1308	EST02338
1324	EST02354
1344	EST02374
1356	EST02386
1365	EST02396
1383	EST02415
1399	EST02433
1401	EST02435
1405	EST02439
1417	EST02452
1451	EST02487
1457	EST02493
1463	EST02500
1473	EST02510
1479	EST02516
1516	EST02555
1528	EST02569
1531	EST02572
1544	EST02586
1551	EST02593
1558	EST02601
1561	EST02604
1581	EST02625
1586	EST02631
1591	EST02636
1616	EST02661
1624	EST02670
1630	EST02676
1637	EST00796
1639	EST00799
1649	EST00808
1651	EST00810
1677	EST00835
1682	EST00839
1694	EST00849
1706	EST00857
1708	EST00858
1710	EST00860
1716	EST00865

SEQ ID#	EST#
1718	EST00867
1731	EST00879
1742	EST00887
1746	EST00891
1760	EST00903
1767	EST00907
1769	EST00909
1777	EST00913
1807	EST00941
1809	EST00943
1820	EST00951
1829	EST00958
1849	EST00975
1860	EST00983
1866	EST00989
1871	EST00994
1888	EST01005
1890	EST01007
1892	EST01009
1903	EST01018
1904	EST01019
1914	EST01026
1930	EST01040
1944	EST01050
1949	EST01054
1962	EST01062
1973	EST01071
1977	EST01075
1982	EST01080
1991	EST01088
1993	EST01090
2000	EST01097
2001	EST01098
2012	EST01106
2013	EST01107
2024	EST01117
2043	EST01131
2051	EST01138
2056	EST01142
2058	EST01144
2059	EST01145
2064	EST01149
2090	EST01167
2094	EST01171
2116	EST01192
2117	EST01193
2128	EST01202
2131	EST01205
2134	EST01208
2144	EST01216
2145	EST01217
2150	EST01222
2155	EST01227
2161	EST01231
2163	EST01238
2174	EST01242
2176	EST01244
2189	EST01255
2214	EST01272
2225	EST01278
2227	EST01279
2233	EST01284
2235	EST01286
2236	EST01287
2255	EST01302
2259	EST01304
2263	EST01307

SEQ ID#	EST#
2267	EST01756
2281	EST01321
2283	EST01322
2300	EST01333
2303	EST01335
2303	EST01335
2314	EST01345
2334	EST01358
2339	EST01362
2342	EST01365
2348	EST01371
2358	EST01379
2367	EST01388
2373	EST01393
2374	EST01394
2393	EST01417
2394	EST01418
2396	EST01420

Table 7: ESTs with Good Probability of Containing Coding Sequence

SEQ ID#	EST#
20	EST00024
72	EST00071
82	EST00078
88	EST00084
137	EST00272
177	EST00328
193	EST00156
200	EST00162
218	EST00175
228	EST00179
247	EST00279
264	EST00204
267	EST00297
296	EST00228
371	EST00426
385	EST00436
392	EST00442
414	EST00460
433	EST00474
453	EST00492
471	EST00505
496	EST00525
524	EST00544
526	EST00546
529	EST00549
549	EST00563
557	EST00569
578	EST00588
596	EST00602
607	EST00610
619	EST00619
657	EST00646
660	EST00649
689	EST00673
695	EST00679
699	EST00682
729	EST00703
742	EST00713
747	EST00717
755	EST00723
759	EST00725
776	EST00738
778	EST00740
782	EST01551
829	EST00768
835	EST00772
836	EST00773
862	EST01872
881	EST01881

SEQ ID#	EST#
884	EST01884
924	EST01926
929	EST01932
938	EST01941
971	EST01985
995	EST02009
996	EST02010
1031	EST02046
1041	EST02057
1083	EST02102
1099	EST02118
1105	EST02124
1113	EST02133
1139	EST02161
1146	EST02168
1196	EST02221
1210	EST02238
1233	EST02262
1285	EST02314
1331	EST02361
1388	EST02421
1418	EST02453
1439	EST02475
1502	EST02540
1537	EST02578
1563	EST02606
1599	EST02644
1602	EST02647
1693	EST00848
1695	EST00850
1729	EST00877
1730	EST00878
1738	EST00883
1739	EST00885
1743	EST00888
1768	EST00908
1780	EST00916
1804	EST00938
1805	EST00939
1811	EST00945
1819	EST00950
1826	EST00956
1830	EST00959
1845	EST00971
1848	EST00974
1853	EST00977
1967	EST01066
1992	EST01089
1994	EST01091

SEQ ID#	EST#
1997	EST01094
2046	EST01134
2101	EST01177
2102	EST01178
2105	EST01181
2106	EST01182
2141	EST01213
2184	EST01251
2196	EST01260
2203	EST01264
2232	EST01283
2308	EST01339
2345	EST01368
2346	EST01369
2351	EST01373
2354	EST01375
2355	EST01376
2359	EST01380
2362	EST01383
2378	EST01397
2399	EST01423
2407	EST02714

Table 8: ESTs with Marginal Probability of Containing Coding Sequence

SEQ ID#	EST#
11	EST00018
12	EST00274
24	EST00027
45	EST00364
79	EST00076
90	EST00302
110	EST00096
144	EST00120
145	EST00121
192	EST00155
222	EST00177
234	EST00184
277	EST00212
319	EST00381
368	EST00423
370	EST00425
387	EST00438
402	EST00451
415	EST00461
418	EST00464
426	EST00470
503	EST00528
517	EST00539
522	EST00543
532	EST00551
540	EST00557
570	EST00580
573	EST00583
576	EST00586
613	EST00615
617	EST00617
626	EST00622
681	EST00665
726	EST00700
727	EST00701
738	EST00711
745	EST00715
752	EST00720
791	EST00746
795	EST00749
803	EST00756
845	EST00777
852	EST00782
854	EST00784
907	EST01907
912	EST01912
935	EST01938
968	EST01981
985	EST01999
988	EST02002
1043	EST02059
1081	EST02100
1089	EST02108
1116	EST02136
1134	EST02154
1205	EST02233
1222	EST02251
1224	EST02253
1228	EST02257
1267	EST02296
1301	EST02331
1397	EST02431
1448	EST02484
1480	EST02517
1493	EST02531
1499	EST02537
1503	EST02541
1527	EST02568
1536	EST02577
1548	EST02590
1562	EST02605
1572	EST02615
1575	EST02618
1595	EST02640
1608	EST02653
1610	EST02655
1621	EST02667
1627	EST02674
1629	EST02677
1631	EST02678
1683	EST00840
1692	EST00847
1751	EST00895
1756	EST00900
1764	EST02690
1770	EST00910
1793	EST00929
1847	EST00973
1877	EST00998
1897	EST01012
1900	EST01015
1939	EST01655
1940	EST01046
1954	EST01058
1990	EST01087
2008	EST01103
2031	EST01123
2041	EST01130
2044	EST01132
2060	EST01146
2100	EST01176
2136	EST01210
2153	EST01225
2204	EST01265
2212	EST01270
2248	EST01297
2250	EST01299
2266	EST01310
2309	EST01340
2347	EST01370
2388	EST01406
2398	EST01422
2405	EST01427

Table 9: ESTs with Poor Coding Probability

SEQ ID#	EST#
1	EST00007
2	EST00009
3	EST00010
4	EST00011
5	EST00012
6	EST00013
8	EST00234
10	EST00016
14	EST00019
16	EST00021
17	EST00022
18	EST00373
19	EST00023
21	EST00025
23	EST00026
25	EST00028
27	EST00029
28	EST00030
29	EST00031
30	EST00032
31	EST00033
32	EST00233
33	EST00034
34	EST00035
35	EST00036
36	EST00037
39	EST00039
40	EST00040
41	EST00041
42	EST00042
46	EST00044
47	EST00046
49	EST00047
50	EST00048
51	EST00049
52	EST00052
53	EST00054
54	EST00055
55	EST00056
56	EST00057
57	EST00058
58	EST00059
59	EST00061
60	EST00062
63	EST00065
64	EST00066
67	EST00351
68	EST00068
69	EST00360
71	EST00070
73	EST00072
74	EST00073
76	EST00075
80	EST00077
81	EST00315
83	EST00079
84	EST00080
85	EST00081
86	EST00082
87	EST00083
89	EST00085
91	EST00086
92	EST00087
94	EST00353
95	EST00088
96	EST00089
99	EST00316

SEQ ID#	EST#
100	EST00090
101	EST00091
103	EST00317
104	EST00354
105	EST00365
107	EST00093
109	EST00095
111	EST00281
112	EST00318
113	EST00097
116	EST00100
117	EST00319
118	EST00101
119	EST00102
120	EST00103
121	EST00104
122	EST00105
123	EST00106
125	EST00108
126	EST00109
127	EST00320
129	EST00321
130	EST00355
131	EST00322
133	EST00111
134	EST00375
135	EST00112
136	EST00113
138	EST00114
139	EST00116
140	EST00117
141	EST00118
142	EST00323
143	EST00119
146	EST00122
147	EST00292
148	EST00236
149	EST00123
150	EST00124
151	EST00125
152	EST00126
153	EST00127
154	EST00128
155	EST00129
157	EST00131
158	EST00132
159	EST00325
160	EST00326
162	EST00133
163	EST00134
165	EST00136
167	EST00138
168	EST00140
169	EST00141
170	EST00295
171	EST00327
172	EST00142
173	EST00143
175	EST00144
178	EST00294
182	EST00329
184	EST00149
185	EST00150
186	EST00151
190	EST00153
191	EST00154
194	EST00157

SEQ ID#	EST#
195	EST00158
196	EST00159
197	EST00160
198	EST00161
199	EST00277
203	EST00164
204	EST00235
206	EST00166
207	EST00167
209	EST00331
210	EST00168
211	EST00332
212	EST00169
213	EST00170
214	EST00171
216	EST00173
219	EST00176
220	EST00372
221	EST00359
224	EST00356
225	EST00178
226	EST00333
229	EST00180
231	EST00334
232	EST00182
233	EST00183
235	EST00185
236	EST00186
237	EST00187
238	EST00188
239	EST00189
240	EST00335
241	EST00191
242	EST00192
243	EST00193
244	EST00194
245	EST00347
246	EST00196
250	EST00197
252	EST00198
254	EST00200
255	EST00201
256	EST00345
257	EST00337
259	EST00202
260	EST00357
261	EST00338
262	EST00339
265	EST00205
266	EST00206
272	EST00340
274	EST00268
275	EST00209
278	EST00342
279	EST00213
280	EST00343
283	EST00215
284	EST00216
286	EST00217
287	EST00218
288	EST00219
289	EST00220
290	EST00221
291	EST00222
292	EST00223
293	EST00224
294	EST00225

SEQ ID#	EST#
295	EST00226
297	EST00230
298	EST00231
302	EST00303
303	EST00348
304	EST00307
305	EST00308
306	EST00309
307	EST00312
308	EST00314
309	EST00174
315	EST00008
316	EST00378
317	EST00379
318	EST00380
320	EST00382
321	EST00383
322	EST00384
323	EST00385
325	EST00386
326	EST00387
327	EST00388
328	EST00389
329	EST00390
330	EST00391
331	EST00392
332	EST00393
334	EST00395
335	EST00396
337	EST00398
340	EST00402
341	EST00403
342	EST00404
344	EST00405
345	EST00406
347	EST01829
348	EST01830
349	EST01831
350	EST00407
351	EST00408
352	EST00409
353	EST00410
354	EST01433
355	EST00411
356	EST00412
357	EST00413
358	EST00414
359	EST00415
360	EST00416
361	EST00417
363	EST00419
364	EST00420
365	EST01434
366	EST00421
367	EST00422
369	EST00424
372	EST00427
373	EST01832
374	EST00428
375	EST00429
376	EST01436
377	EST00430
378	EST00431
379	EST00432
380	EST01439
381	EST00433
382	EST00434

SEQ ID#	EST#
383	EST00435
384	EST01440
386	EST00437
388	EST00439
390	EST01442
391	EST00441
393	EST00443
395	EST00445
397	EST00446
398	EST00447
399	EST00448
400	EST00449
401	EST00450
403	EST00452
404	EST00453
405	EST00454
406	EST00455
407	EST00456
408	EST00457
409	EST01444
410	EST00458
411	EST00459
412	EST01445
416	EST00462
417	EST00463
419	EST00465
420	EST00466
421	EST00467
422	EST01447
423	EST00468
424	EST01448
425	EST00469
427	EST01449
428	EST01451
429	EST00471
431	EST00473
432	EST01452
434	EST00475
435	EST00476
436	EST00477
437	EST00478
438	EST00479
439	EST00480
440	EST01454
442	EST01456
443	EST00482
444	EST00483
446	EST00485
447	EST00486
448	EST00487
449	EST00488
450	EST00489
451	EST00490
452	EST00491
455	EST00494
457	EST00495
458	EST00496
459	EST00497
460	EST01457
461	EST01836
462	EST00498
464	EST00499
465	EST00500
466	EST00501
467	EST00502
468	EST00503
470	EST00504

SEQ ID#	EST#
473	EST00506
474	EST00507
477	EST01463
478	EST00510
479	EST00511
480	EST01464
481	EST00512
482	EST01465
483	EST00513
484	EST00514
487	EST00516
488	EST00517
489	EST00518
490	EST00519
491	EST00520
492	EST00521
495	EST00524
497	EST00526

498	EST01467
499	EST01468
500	EST00527
501	EST02715
502	EST01469
507	EST00530
508	EST00531
509	EST01472
510	EST00532
511	EST00533
512	EST00534
513	EST00535
514	EST00536
515	EST00537
519	EST00541
520	EST00542
521	EST01474
523	EST01838
525	EST00545
527	EST00547
528	EST00548
530	EST01477
531	EST00550
533	EST00552
534	EST01478
535	EST00553
536	EST01479
537	EST00554
538	EST00555
539	EST00556
541	EST00558
542	EST01480
543	EST00559
544	EST00560
545	EST01481
547	EST00561
548	EST00562
550	EST00564
553	EST00566
555	EST01483
556	EST00568
558	EST01484
560	EST01485
561	EST00571
562	EST00572
563	EST00573
564	EST00574
565	EST00575
566	EST00576
567	EST00577
568	EST00578
569	EST00579

SEQ ID#	EST#
571	EST00581
572	EST00582
574	EST00584
575	EST00585
577	EST00587
580	EST00590
581	EST00591
583	EST00593
584	EST00594
585	EST00595
586	EST00596
587	EST01488
588	EST00597
589	EST00598
590	EST00599
591	EST01489
592	EST00600
593	EST00601
595	EST01840
597	EST00603
598	EST00604
599	EST00605

600	EST01492
601	EST01493
603	EST01494
604	EST00607
605	EST00608
609	EST01496
610	EST00612
611	EST00613
612	EST00614
615	EST00616
616	EST01497
618	EST01498
620	EST01499
622	EST01843
623	EST00621
624	EST01500
625	EST01844
627	EST00623
628	EST01503
629	EST00624
630	EST01505
631	EST00625
632	EST00626
633	EST00627
634	EST00628
636	EST01507
637	EST00630
638	EST00631
640	EST01509
641	EST00633
643	EST00635
645	EST00637
646	EST00638
647	EST00639
648	EST00640
649	EST00641
651	EST00643
652	EST01510
654	EST00644
655	EST00645
656	EST01513
658	EST00647
659	EST00648
661	EST00650
662	EST00651
663	EST00652
664	EST00653
665	EST00654

SEQ ID#	EST#
666	EST01514
667	EST00655
668	EST00656
669	EST08657
670	EST00658
671	EST00659
672	EST00660
673	EST01515
674	EST01516
675	EST00661
676	EST00662
677	EST00663
678	EST01517
679	EST01518
680	EST00664
682	EST00666
683	EST00667
684	EST00668
685	EST00669
686	EST00670
688	EST00672
690	EST00674
692	EST00676
693	EST00677
694	EST00678
696	EST01521

697	EST00680
698	EST00681
701	EST01522
702	EST00684
703	EST00685
704	EST00686
705	EST00687
706	EST00688
708	EST00689
709	EST00690
710	EST00691
711	EST00692
712	EST00693
713	EST00694
714	EST00695
715	EST01523
716	EST01524
717	EST01525
718	EST00696
719	EST01526
720	EST00697
721	EST01527
722	EST01528
723	EST00698
725	EST00699
728	EST00702
730	EST00704
731	EST00705
732	EST00706
733	EST00707
734	EST00708
735	EST00709
736	EST01532
737	EST00710
739	EST01534
740	EST01535
741	EST00712
744	EST01537
746	EST00716
748	EST01850
749	EST00719
750	EST01539
751	EST01540
754	EST00722

SEQ ID#	EST#
756	EST01541
758	EST00724
761	EST01544
762	EST00727
763	EST00728
765	EST00730
766	EST00731
767	EST00732
768	EST00733
770	EST00735
771	EST01546
772	EST00736
774	EST01548
775	EST00737
777	EST00739
779	EST00741
780	EST01549
781	EST01550
783	EST01552
785	EST01553
786	EST00742
787	EST00743
788	EST00744
789	EST00745
790	EST01554
792	EST00747
793	EST00748
794	EST01555
796	EST00750
797	EST00751

799	EST00752
800	EST00753
801	EST00754
804	EST00757
805	EST00758
806	EST00759
807	EST00760
809	EST00762
810	EST00763
811	EST00764
813	EST00765
814	EST00766
815	EST01855
816	EST01856
817	EST01857
818	EST01858
819	EST01859
820	EST01860
822	EST01863
825	EST01866
826	EST01867
827	EST01558
828	EST00767
830	EST01559
831	EST00769
832	EST00770
837	EST01561
838	EST00774
839	EST01562
840	EST00775
841	EST00776
842	EST01563
843	EST01564
844	EST01565
846	EST00778
847	EST00779
848	EST01566
849	EST01567
850	EST00780
851	EST00781

SEQ ID#	EST#
853	EST00783
855	EST00785
856	EST01568
857	EST01868
858	EST01869
859	EST01870
860	EST00786
861	EST01871
863	EST01873
864	EST00787
865	EST01569
866	EST01874
867	EST01875
868	EST01876
869	EST00788
870	EST00789
871	EST00790
872	EST00791
873	EST00792
874	EST00793
875	EST00794
876	EST00795
877	EST01877
878	EST01878
879	EST01879
880	EST01880
882	EST01882
883	EST01883
885	EST01885
887	EST01887
889	EST01889
890	EST01890
892	EST01892
893	EST01893

SEQ ID#	EST#
894	EST01894
895	EST01895
896	EST01896
897	EST01897
898	EST01898
899	EST01899
900	EST01900
901	EST01901
902	EST01902
903	EST01903
904	EST01904
905	EST01905
906	EST01906
908	EST01908
909	EST01909
910	EST01910
911	EST01911
914	EST01914
915	EST01915
916	EST01917
917	EST01919
918	EST01920
920	EST01922
921	EST01923
922	EST01924
923	EST01925
925	EST01927
926	EST01929
927	EST01930
928	EST01931
931	EST01934
932	EST01935
933	EST01936
934	EST01937
937	EST01940
939	EST01943
940	EST01944
941	EST01945
942	EST01947
943	EST01948
944	EST01949
945	EST01950
946	EST01953
947	EST01954
949	EST01958
950	EST01959
953	EST01962
954	EST01963
956	EST01968
957	EST01969
958	EST01970
959	EST01972
960	EST01973
961	EST01974
962	EST01975
963	EST01976
964	EST01977
966	EST01979
967	EST01980
970	EST01983
972	EST01986
974	EST01988
975	EST01989
976	EST01990
977	EST01991
978	EST01992
981	EST01995
982	EST01996
983	EST01997
984	EST01998
987	EST02001
989	EST02003
990	EST02004
991	EST02005

992	EST02006
994	EST02008
997	EST02011
999	EST02013
1001	EST02015
1002	EST02016
1003	EST02017
1005	EST02019
1006	EST02020
1008	EST02022
1009	EST02023
1010	EST02024
1011	EST02025
1012	EST02026
1013	EST02027
1014	EST02028
1015	EST02029
1016	EST02030
1017	EST02031
1019	EST02033
1022	EST02036
1023	EST02037
1024	EST02038
1025	EST02040
1026	EST02041
1027	EST02042
1028	EST02043
1029	EST02044
1030	EST02045
1032	EST02048
1033	EST02049
1036	EST02052

SEQ ID#	EST#
1037	EST02053
1038	EST02054
1040	EST02056
1042	EST02058
1044	EST02060
1045	EST02061
1046	EST02062
1048	EST02064
1049	EST02065
1050	EST02066
1051	EST02067
1052	EST02068
1053	EST02069
1054	EST02070
1055	EST02071
1056	EST02072
1057	EST02073
1058	EST02074
1059	EST02075
1060	EST02076
1061	EST02078
1062	EST02079
1063	EST02081
1064	EST02082
1065	EST02083
1066	EST02084
1067	EST02085
1068	EST02086
1070	EST02088
1071	EST02089
1072	EST02090
1073	EST02091
1074	EST02092
1075	EST02093
1076	EST02094
1077	EST02096
1078	EST02097
1079	EST02098
1080	EST02099
1082	EST02101
1084	EST02103
1085	EST02104
1086	EST02105
1087	EST02106
1088	EST02107
1091	EST02110
1093	EST02112
1095	EST02114
1097	EST02116
1098	EST02117
1100	EST02119
1101	EST02120
1102	EST02121
1104	EST02123
1106	EST02125
1107	EST02126
1108	EST02127
1109	EST02128
1110	EST02129
1111	EST02131
1112	EST02132
1114	EST02134
1117	EST02137
1119	EST02139
1120	EST02140
1121	EST02141
1122	EST02142
1123	EST02143
1124	EST02144
1125	EST02145

SEQ ID#	EST#
1127	EST02147
1128	EST02148
1130	EST02150
1131	EST02151
1132	EST02152
1135	EST02155
1136	EST02156
1137	EST02157
1138	EST02159
1140	EST02162
1142	EST02164
1143	EST02165
1144	EST02166
1145	EST02167
1148	EST02170
1149	EST02171
1150	EST02172
1152	EST02174
1153	EST02175
1154	EST02176
1155	EST02177
1156	EST02178
1157	EST02180
1158	EST02181
1159	EST02182
1160	EST02183
1161	EST02184
1162	EST02185
1164	EST02188
1165	EST02189
1166	EST02190
1167	EST02191
1168	EST02193
1169	EST02194
1170	EST02195
1171	EST02196
1172	EST02197
1173	EST02198
1174	EST02199
1175	EST02200
1176	EST02201
1177	EST02202
1178	EST02203
1179	EST02204
1180	EST02205
1182	EST02207
1184	EST02209
1185	EST02210
1186	EST02211
1187	EST02212
1188	EST02213
1189	EST02214
1190	EST02215
1191	EST02216
1192	EST02217
1193	EST02218
1194	EST02219
1195	EST02220
1197	EST02222
1198	EST02223
1199	EST02224
1200	EST02226
1201	EST02228
1202	EST02229
1203	EST02230
1204	EST02232
1206	EST02234
1207	EST02235
1208	EST02236
1209	EST02237

SEQ ID#	EST#
1211	EST02239
1212	EST02240
1213	EST02241
1214	EST02242
1215	EST02244
1216	EST02245
1217	EST02246
1218	EST02247
1219	EST02248
1220	EST02249
1221	EST02250
1223	EST02252
1225	EST02254
1226	EST02255
1227	EST02256
1232	EST02261
1234	EST02263
1235	EST02264
1236	EST02265
1237	EST02266
1238	EST02267
1239	EST02268
1240	EST02269
1241	EST02270
1242	EST02271
1244	EST02273
1246	EST02275
1247	EST02276
1248	EST02277
1249	EST02278
1250	EST02279
1251	EST02280
1252	EST02281
1253	EST02282
1254	EST02283
1255	EST02284
1256	EST02285
1257	EST02286
1258	EST02287
1259	EST02288
1260	EST02289
1261	EST02290
1262	EST02291
1263	EST02292
1268	EST02297
1269	EST02298
1270	EST02299
1271	EST02300
1272	EST02301
1273	EST02302
1274	EST02303
1275	EST02304
1276	EST02305
1278	EST02307
1279	EST02308
1280	EST02309
1281	EST02310
1282	EST02311
1283	EST02312
1284	EST02313
1286	EST02316
1288	EST02318
1289	EST02319
1290	EST02320
1291	EST02321
1292	EST02322
1293	EST02323
1294	EST02324
1295	EST02325
1296	EST02326

SEQ ID#	EST#
1298	EST02328
1299	EST02329
1300	EST02330
1302	EST02332
1303	EST02333
1304	EST02334
1305	EST02335
1306	EST02336
1307	EST02337
1309	EST02339
1310	EST02340
1311	EST02341
1313	EST02343
1314	EST02344
1315	EST02345
1316	EST02346
1317	EST02347
1318	EST02348
1319	EST02349
1320	EST02350
1321	EST02351
1322	EST02352
1323	EST02353
1325	EST02355
1326	EST02356
1327	EST02357
1328	EST02358
1329	EST02359
1330	EST02360
1333	EST02363
1334	EST02364
1335	EST02365
1336	EST02366
1337	EST02367
1338	EST02368
1339	EST02369
1342	EST02372
1343	EST02373
1345	EST02375
1346	EST02376
1347	EST02377
1349	EST02379
1350	EST02380
1351	EST02381
1352	EST02382
1353	EST02383
1354	EST02384
1355	EST02385
1357	EST02387
1358	EST02388
1359	EST02390
1360	EST02391
1361	EST02392
1362	EST02393
1363	EST02394
1364	EST02395
1366	EST02397
1367	EST02398
1368	EST02399
1370	EST02401
1372	EST02403
1373	EST02404
1375	EST02406
1376	EST02407
1377	EST02408
1378	EST02409
1379	EST02410
1380	EST02411
1381	EST02413
1382	EST02414

SEQ ID#	EST#
1384	EST02416
1387	EST02419
1389	EST02422
1390	EST02423
1391	EST02424
1392	EST02425
1393	EST02426
1394	EST02427
1396	EST02430
1398	EST02432
1400	EST02434
1402	EST02436
1403	EST02437
1404	EST02438
1406	EST02440
1407	EST02441
1410	EST02444
1411	EST02445
1414	EST02448
1415	EST02449
1416	EST02450
1419	EST02454
1420	EST02456
1421	EST02457
1422	EST02458
1423	EST02459
1424	EST02460
1425	EST02461
1426	EST02462
1428	EST02464
1429	EST02465
1431	EST02467
1432	EST02468
1433	EST02469
1434	EST02470
1435	EST02471
1436	EST02472
1437	EST02473
1438	EST02474
1440	EST02476
1442	EST02478
1443	EST02479
1444	EST02480
1445	EST02481
1446	EST02482
1447	EST02483
1450	EST02486
1452	EST02488
1453	EST02489
1454	EST02490
1455	EST02491
1456	EST02492
1458	EST02495
1459	EST02496
1460	EST02497
1461	EST02498
1462	EST02499
1464	EST02501
1466	EST02503
1467	EST02504
1469	EST02506
1470	EST02507
1471	EST02508
1472	EST02509
1474	EST02511
1475	EST02512
1476	EST02513
1477	EST02514
1481	EST02518
1482	EST02519
SEQ ID#	EST#
1483	EST02520
1484	EST02521
1485	EST02522
1486	EST02523
1487	EST02524
1488	EST02525
1489	EST02526
1490	EST02527
1491	EST02529
1494	EST02532
1497	EST02535
1498	EST02536
1501	EST02539
1504	EST02542
1506	EST02545
1507	EST02546
1508	EST02547
1509	EST02548
1510	EST02549
1512	EST02551
1513	EST02552
1514	EST02553
1515	EST02554
1517	EST02558
1518	EST02559
1519	EST02560
1520	EST02561
1521	EST02562
1522	EST02563
1523	EST02564
1524	EST02565
1525	EST02566
1526	EST02567
1529	EST02570
1530	EST02571
1532	EST02573
1533	EST02574
1534	EST02575
1535	EST02576
1538	EST02579
1539	EST02580
1540	EST02581
1541	EST02582
1542	EST02583
1545	EST02587
1546	EST02588
1547	EST02589
1549	EST02591
1550	EST02592
1552	EST02594
1553	EST02595
1554	EST02597
1555	EST02598
1556	EST02599
1557	EST02600
1559	EST02602
1560	EST02603
1564	EST02607
1565	EST02608
1568	EST02611
1569	EST02612
1570	EST02613
1571	EST02614
1573	EST02616
1574	EST02617
1576	EST02619
1577	EST02620
1578	EST02621
1579	EST02622
1580	EST02623
SEQ ID#	EST#
1582	EST02626
1583	EST02628
1584	EST02629
1585	EST02630
1587	EST02632
1590	EST02635
1592	EST02637
1593	EST02638
1594	EST02639
1596	EST02641
1597	EST02642
1598	EST02643
1600	EST02645
1601	EST02646
1603	EST02648
1604	EST02649
1605	EST02650
1606	EST02651
1607	EST02652
1609	EST02654
1611	EST02656
1612	EST02657
1613	EST02658
1614	EST02659
1615	EST02660
1617	EST02662
1618	EST02663
1619	EST02665
1620	EST02666
1622	EST02668
1623	EST02669
1625	EST02672
1626	EST02673
1628	EST02675
1632	EST02679
1633	EST02680
1634	EST02681
1635	EST02682
1636	EST02684
1638	EST00798
1640	EST00800
1641	EST00801
1642	EST00802
1643	EST00803
1645	EST00804
1646	EST00805
1647	EST00806
1648	EST00807
1650	EST00809
1652	EST00811
1653	EST00812
1655	EST00813
1656	EST00814
1657	EST00815
1658	EST00816
1659	EST00817
1660	EST00818
1661	EST00819
1662	EST00820
1663	EST00821
1664	EST00822
1665	EST00823
1666	EST00824
1668	EST00826
1669	EST00827
1670	EST00828
1671	EST00829
1672	EST00830
1673	EST00831
1674	EST00832
SEQ ID#	EST#
1675	EST00833
1676	EST00834
1678	EST00836
1679	EST00837
1680	EST00838
1684	EST00841
1685	EST00842
1686	EST01574
1687	EST00843
1688	EST00844
1689	EST00845
1690	EST00846
1691	EST01577
1696	EST00851
1697	EST00852
1702	EST00854
1703	EST00855
1705	EST00856
1707	EST01581
1709	EST00859
1711	EST00861
1712	EST00862
1713	EST00863
1714	EST00864
1717	EST00866
1719	EST00868
1720	EST00869
1721	EST00870
1722	EST00871
1723	EST00872
1724	EST00873
1725	EST00874
1727	EST00875
1728	EST00876
1732	EST01590
1733	EST01591
1734	EST00880
1735	EST00881
1736	EST01592
1737	EST00882
1740	EST02687
1741	EST00886
1744	EST00889
1745	EST00890
1747	EST00892
1748	EST00893
1749	EST01593
1750	EST00894
1752	EST00896
1753	EST00897
1754	EST00898
1755	EST00899
1757	EST01594
1758	EST00901
1759	EST00902
1761	EST01598
1762	EST00904
1763	EST00905
1765	EST01600
1766	EST00906
1772	EST02691
1773	EST00911
1774	EST00912
1775	EST02692
1776	EST01603
1778	EST00914
1779	EST00915
1781	EST00917
1782	EST00918
1783	EST00919
SEQ ID#	EST#
1784	EST00920
1785	EST00921
1786	EST00922
1787	EST00923
1788	EST00924
1789	EST00925
1790	EST00926
1791	EST00927
1792	EST00928
1794	EST01607
1795	EST00930
1796	EST00931
1797	EST00932
1798	EST00933
1799	EST00934
1800	EST00935
1801	EST00936
1802	EST00937
1803	EST01613
1806	EST00940
1808	EST00942
1810	EST00944
1812	EST02693
1813	EST00946
1814	EST00947
1815	EST01615
1816	EST00948
1817	EST00949
1818	EST01616
1821	EST00952
1822	EST00953
1823	EST00954
1824	EST01617
1825	EST00955
1827	EST01618
1828	EST00957
1831	EST01619
1832	EST00960
1833	EST00961
1835	EST00962
1836	EST01622
1837	EST00963
1838	EST00964
1839	EST00965
1840	EST00966
1841	EST00967
1842	EST00968
1843	EST00969
1844	EST00970
1846	EST00972
1850	EST01624
1851	EST00976
1854	EST00978
1855	EST00979
1857	EST00980
1858	EST00981
1859	EST00982
1861	EST00984
1862	EST00985
1863	EST00986
1864	EST00987
1865	EST00988
1867	EST00990
1868	EST00991
1870	EST00993
1872	EST00995
1873	EST01630
1874	EST00996
1875	EST01631
1876	EST00997
SEQ ID#	EST#
1878	EST00999
1879	EST01633
1881	EST01000
1882	EST01638
1883	EST01001
1884	EST01002
1886	EST01003
1887	EST01004
1889	EST01006
1891	EST01008
1893	EST01642
1895	EST01010
1898	EST01013
1899	EST01014
1901	EST01016
1902	EST01017
1905	EST01020
1906	EST01021

1907	EST01022
1908	EST01023
1909	EST01024
1911	EST02694
1912	EST01025
1913	EST01646
1915	EST01027
1916	EST01028
1917	EST01029
1918	EST02695
1919	EST01030
1920	EST01031
1921	EST01647
1922	EST01032
1923	EST01033
1924	EST01034
1925	EST01035
1926	EST01036
1927	EST01037
1929	EST01039
1932	EST01042
1934	EST01043
1935	EST01044
1936	EST01045
1937	EST01652
1938	EST01654
1941	EST01047
1942	EST01048
1943	EST01049
1945	EST01051
1946	EST02696
1947	EST01052
1948	EST01053
1950	EST01055
1951	EST01056
1952	EST01057
1955	EST01662
1957	EST01059
1958	EST01060
1959	EST01061
1963	EST01063
1964	EST01064
1966	EST01065
1968	EST01067
1969	EST01068
1970	EST01666
1971	EST01069
1972	EST01070
1975	EST01073
1976	EST01074
1978	EST01076
1979	EST01077

SEQ ID#	EST#
1980	EST01078
1981	EST01079
1983	EST01081
1984	EST01082
1985	EST01083
1986	EST01084
1988	EST01085
1989	EST01086
1995	EST01092
1996	EST01093
1998	EST01095
1999	EST01096
2002	EST01099
2003	EST01675
2005	EST01100
2006	EST01101
2007	EST01102
2009	EST01677
2010	EST01104
2011	EST01105
2014	EST01108
2015	EST01109
2016	EST01110
2018	EST01111
2019	EST01112
2020	EST01113
2021	EST01114
2022	EST01115
2023	EST01116
2025	EST01118
2026	EST01119
2027	EST01120
2028	EST01121
2029	EST01682
2030	EST01122
2033	EST01684
2034	EST01124
2035	EST01125
2036	EST01126
2037	EST01686
2038	EST01127
2039	EST01128
2040	EST01129
2042	EST01688
2045	EST01133
2047	EST01135
2048	EST01136
2049	EST01689
2050	EST01137
2052	EST01139
2053	EST01140
2054	EST01141
2055	EST01690
2057	EST01143
2061	EST01147
2062	EST02701
2063	EST01148
2065	EST01691
2066	EST01692
2067	EST01693
2069	EST01150
2070	EST01151
2072	EST01152
2074	EST01698
2075	EST01153
2076	EST02702
2077	EST01154
2078	EST01155
2079	EST01156
2080	EST01157

SEQ ID#	EST#
2081	EST01158
2082	EST01159
2083	EST01160
2084	EST01161
2085	EST01162
2086	EST01163
2087	EST01164
2088	EST01166
2091	EST01168
2093	EST01170
2095	EST01701
2096	EST01172
2097	EST01173
2098	EST01174
2099	EST01175
2103	EST01179
2104	EST01180
2107	EST01183
2108	EST01184
2109	EST01185
2110	EST01186
2111	EST01187
2112	EST01188
2113	EST01189
2114	EST01190
2115	EST01191
2118	EST01194
2119	EST01195
2122	EST01197
2123	EST01713
2124	EST01198
2125	EST01199
2126	EST01200
2127	EST01201
2129	EST01203
2130	EST01204
2132	EST01206
2133	EST01207
2135	EST01209
2137	EST01211
2139	EST01716
2140	EST01212
2142	EST01214
2143	EST01215
2147	EST01219
2148	EST01220
2151	EST01223
2152	EST01224
2154	EST01226
2156	EST01718
2157	EST01719
2158	EST01228
2159	EST01229
2160	EST01230
2162	EST01232
2163	EST01233
2164	EST01234
2165	EST01720
2166	EST01236
2167	EST01237
2169	EST01722
2170	EST01239
2171	EST01240
2172	EST01241
2175	EST01243
2177	EST01245
2178	EST01726
2179	EST01246
2180	EST01247
2181	EST01248

SEQ ID#	EST#
2182	EST01249
2183	EST01250
2185	EST01252
2186	EST01253
2187	EST01727
2188	EST01254
2190	EST01728
2191	EST01256
2193	EST01258
2194	EST01729
2195	EST01259
2197	EST01261
2198	EST01730
2199	EST01262
2200	EST01731
2201	EST01263
2202	EST01732
2205	EST01735
2206	EST01736
2208	EST01267
2209	EST02717
2210	EST01268
2211	EST01269
2213	EST01271
2215	EST01273
2218	EST01274
2219	EST01275
2220	EST01740
2221	EST01741
2222	EST01276
2223	EST01742
2224	EST01277
2228	EST01280
2229	EST01281
2231	EST01746
2237	EST01288
2238	EST01289
2239	EST01290
2240	EST01291
2241	EST01747
2242	EST01292
2243	EST01293
2244	EST01294
2246	EST01295
2247	EST01296
2249	EST01298
2251	EST01300
2252	EST01750
2253	EST01301
2256	EST02718
2257	EST01303
2258	EST01754
2260	EST01305
2261	EST01755
2262	EST01306
2264	EST01308
2265	EST01309
2268	EST01311
2269	EST01312
2270	EST01313
2271	EST01314
2272	EST01762
2273	EST01315
2275	EST01316
2276	EST01317
2277	EST01318
2278	EST01319
2279	EST01320
2280	EST01763
2284	EST01323

SEQ ID#	EST#
2285	EST01768
2287	EST01770
2288	EST01324
2290	EST01772
2291	EST01773
2292	EST01326
2293	EST01327
2294	EST01328
2295	EST01329
2296	EST01330
2298	EST01331
2299	EST01332
2301	EST01334
2304	EST01780
2305	EST01336
2306	EST01337
2310	EST01341
2311	EST01342
2312	EST01343
2313	EST01344
2315	EST01346
2316	EST01782
2317	EST01347
2318	EST01348
2319	EST01349
2321	EST01350
2322	EST01351
2323	EST01789
2325	EST01353
2327	EST01354
2328	EST01355
2329	EST01792
2330	EST01793
2331	EST01356
2332	EST01794
2333	EST01357
2335	EST01359
2336	EST01360
2337	EST01361
2340	EST01802
2341	EST01364
2343	EST01366
2344	EST01367
2349	EST01372
2350	EST02708
2352	EST01374
2356	EST01377
2357	EST01378
2360	EST01381
2361	EST01382
2363	EST01384
2364	EST01385
2365	EST01386
2366	EST01387
2369	EST01811
2370	EST01390
2371	EST01391
2372	EST01392
2375	EST01815
2376	EST01395
2377	EST01396
2379	EST01398
2380	EST01399
2381	EST01400
2382	EST01401
2383	EST01402
2384	EST01403
2385	EST01816
2386	EST01404
2387	EST01405

SEQ ID#	EST#
2389	EST01407
2391	EST01415
2392	EST01416
2395	EST01419
2397	EST01421
2401	EST01424
2403	EST01425
2404	EST01426
2406	EST02713
2409	EST00273

EXAMPLE 10

Functional Groupings of ESTs and Corresponding Genes

By matching new human ESTs to known sequences from other species, the apparent function of the gene corresponding to the EST can be ascertained. The data generated in Example 3 and 4 have been used to categorize 127 of the ESTs of the present invention, and their corresponding genes, into predicted functional groups. (These 127 are ESTs with database matches to sequences from other species for which a function was known.) Two different grouping schemes have been used.

The first scheme separates the sequences into three broad categories: metabolic; regulatory; and structural. These groupings are set out in Table 10.

The second grouping scheme separates the sequences into 13 specific categories: cell surface proteins; developmental control; energy metabolism; kinases and phosphatases; oncogenes; other metabolism-related polypeptides; peptidases and peptidase inhibitors; receptors; structural and cytoskeletal; signal transduction; transporters; transcription, translation, and subcellular localization; and transcription factors. These groupings are set out in Table 11.

Table 10: Three-Class Functional Groupings of ESTs

SEQ ID	EST#	Group	Putative Identification
1834	EST01620	M	AMP deaminase, brain
97	EST00289	M	Aconitase
691	EST00675	M	Alcohol dehydrogenase
2092	EST01700	M	Anion exchanger homolog AE3
396	EST01443	M	CDPdiacylglycerol-serine O-phosphatidyltransfera
1956	EST01663	M	Ca2+-transporting ATPase 2
1039	EST02055	M	Calcium channel
2192	EST01257	M	Diacylglycerol kinase, lymphyocyte
1441	EST02477	M	Diamine acetyltransferase
2289	EST01325	M	Fatty acid synthase
310	EST00377	M	Fo ATPase beta subunit, mitochondrial
1667	EST00825	M	Gamma-aminobutyric acid transporter
1412	EST02446	M	Glutamate-aspartate carrier protein
1020	EST02034	M	Glutaminase
2326	EST01791	M	Inositol-1,4,5-trisphosphate 3-kinase
2173	EST01724	M	Lon protease
1427	EST02463	M	Long-chain-fatty-acid-CoA ligase
2226	EST01744	M	NAD(P)+ transhydrogenase (B-specific)
1566	EST02609	M	Neutrophil oxidase factor
1681	EST01573	M	Nucleoside diphosphate kinase
2254	EST01751	M	Phosphatidylinositol-4,5-bisphosphate phosphodie
93	EST00287	M	Processing enhancing protein
2297	EST01775	M	Prohormone cleavage enzyme
9	EST00376	M	Prolyl endopeptidase
1654	EST01572	M	Protochlorophyllide reductase
38	EST00374	M	RNA polymerase II 6th subunit (RPO26)
1715	EST01583	M	Ribosomal protein L18a
1856	EST01627	M	Ribosomal protein L1a
1974	EST01667	M	Ribosomal protein L3
301	EST00300	M	Ribosomal protein L30
22	EST00301	M	Ribosomal protein S10
2402	EST01826	M	Ribosomal protein S10
463	EST01459	M	Ribosomal protein YL10
2073	EST01697	M	Succinate dehydrogenase flavoprotein
2138	EST01715	M	Succinate dehydrogenase flavoprotein
1771	EST01601	M	Thiosulfate sulfurtransferase (rhodanese)
2121	EST01711	M	Valine-tRNA ligase
1726	EST01588	M	XPR2 alkaline extracellular protease
913	EST01913	M	Clathrin coat assembly protein AP50 homolog
1035	EST02051	M	J1 protein
969	EST01982	R	ADP-ribosylation factor 1
1126	EST02146	R	Calbindin D28
1910	EST01645	R	Calmodulin
485	EST01466	R	Calmodulin-dependent protein kinase, type II, be
2302	EST01779	R	Discs-large tumor suppressor
188	EST00256	R	Enhancer of split
1229	EST02258	R	KUP protein
993	EST02007	R	Kinase 5 protein
2282	EST01764	R	Lamin B receptor
SEQ ID	EST#	Group	Putative Identification
161	EST00247	R	MARCKS (myristoylated alanine-rich protein kinas
769	EST00734	R	MARCKS homolog
1386	EST02418	R	MARCKS homolog
227	EST00259	R	Notch/Xotch
952	EST01961	R	Notch/Xotch
1395	EST02429	R	Nuclear factor 1-like protein (NF1)
2353	EST01806	R	Prohibitin
1069	EST02087	R	Protein kinase C, zeta
1933	EST01650	R	Protein phosphatase 2A beta subunit

202	EST00298	R	Protein-tyrosine phosphatase LRP
1478	EST02515	R	Rab5
1408	EST02442	R	Seven in absentia
300	EST00232	R	Transforming protein (dbl)
1147	EST02169	R	Tyrosine kinase
1348	EST02378	R	cAMP-dependent protein kinase inhibitor
1931	EST01041	R	cAMP-regulated phosphoprotein
1413	EST02447	R	cAMP-specific phosphodiesterase
37	EST00038	R	ras p21-like small GTP-binding protein (smg GDS)
102	EST00248	R	rho H12/ ARH12
299	EST00249	R	smg p25A GDP dissociation inhibitor
189	EST00282	R	trkB
1332	EST02362	R	GA binding protein, beta subunit
1277	EST02306	R	Bib protein
43	EST00371	R	Maternal G10 protein
1704	EST01580	R	Myeloid differentiation primary response gene My
346	EST01828	R	Otd homeotic protein
187	EST00152	R	Wilm's tumor-related protein
249	EST00275	R	Zinc Finger Proteins
413	EST01446	R	Zinc Finger Proteins
469	EST01460	R	Zinc Finger Proteins
833	EST01560	R	Zinc Finger Proteins
1230	EST02259	R	Zinc finger proteins
1496	EST02534	R	Zinc finger proteins
2324	EST01352	R	Zinc Finger Proteins
208	EST00250	S	60K filarial antigen
2320	EST01784	S	60K filarial antigen
251	EST00370	S	Actin, other
2146	EST01218	S	Actin, other
248	EST00271	S	Actinin, alpha
891	EST01891	S	Actinin, alpha
1500	EST02538	S	Actinin, alpha
132	EST00110	S	Agrin
1852	EST01625	S	Agrin
1965	EST01664	S	Amyloid A4
2068	EST01694	S	Amyloid A4
2408	EST00244	S	Amyloid A4
1880	EST01634	S	Axonal glycoprotein TAG-1
2004	EST01676	S	Cofilin
650	EST00642	S	Dilute (myosin heavy chain)
2217	EST01738	S	Gelation factor ABP-280
1885	EST01639	S	Histocompatibility antigen modifier 1
77	EST00257	S	Kinesin
SEQ ID	EST#	Group	Putative Identification
78	EST00258	S	Kinesin
2245	EST01748	S	Kinesin
313	EST00276	S	Lysosomal membrane glycoprotein 1 (LAMP-1)
223	EST00368	S	Microtubule-associated protein 1B
824	EST01865	S	Microtubule-associated protein 1B
2032	EST01683	S	Microtubule-associated protein 1B
2017	EST01678	S	Milk fat globule membrane protein
1567	EST02610	S	Neural cell adhesion molecule L1
506	EST01471	S	Neuraxin
2368	EST01389	S	Radial spoke protein 3
951	EST01960	S	Spectrin, beta
2089	EST01699	S	Sperm membrane protein
653	EST01512	S	Tubulin, alpha
311	EST00270	S	Tubulin, beta
594	EST01490	S	Tubulin, beta
757	EST01542	S	Tubulin, beta
1245	EST02274	S	Tubulin, beta
1589	EST02634	S	Tubulin, beta
1468	EST02505	S	Matrin 3

1371	EST02402	S	Talin
1701	EST00853	S	Unc-104

Group Key: M: Metabolic, R: Regulatory, S: Structural

Table 11: Thirteen-Class Functional Groupings of ESTs

SEQ ID	EST#	Group	Putative Identification
208	EST00250	CS	60K filarial antigen
2320	EST01784	CS	60K filarial antigen
1965	EST01664	CS	Amyloid A4
2068	EST01694	CS	Amyloid A4
2408	EST00244	CS	Amyloid A4
1880	EST01634	CS	Axonal glycoprotein TAG-1
1885	EST01639	CS	Histocompatibility antigen modifier 1
313	EST00276	CS	Lysosomal membrane glycoprotein 1 (LAMP-1)
2017	EST01678	CS	Milk fat globule membrane protein
1567	EST02610	CS	Neural cell adhesion molecule L1
2368	EST01389	CS	Radial spoke protein 3
2089	EST01699	CS	Sperm membrane protein
1277	EST02306	DC	Bib protein
188	EST00256	DC	Enhancer of split
43	EST00371	DC	Maternal G10 protein
1704	EST01580	DC	Myeloid differentiation primary response gene MyD1
227	EST00259	DC	Notch/Xotch
952	EST01961	DC	Notch/Xotch
346	EST01828	DC	Orthodentical homeotic protein
1408	EST02442	DC	Seven in absentia
97	EST00289	EM	Aconitase
310	EST00377	EM	Fo ATPase beta subunit, mitochondrial
485	EST01466	KP	Calmodulin-dependent protein kinase, type II, beta
993	EST02007	KP	Kinase 5 protein
1069	EST02087	KP	Protein kinase C, zeta
1933	EST01650	KP	Protein phosphatase 2A beta subunit
202	EST00298	KP	Protein-tyrosine phosphatase LRP
1348	EST02378	KP	cAMP-dependent protein kinase inhibitor
2302	EST01779	OG	Discs-large tumor suppressor
2353	EST01806	OG	Prohibitin
1478	EST02515	OG	Rab5
300	EST00232	OG	Transforming protein (dbl)
37	EST00038	OG	ras p21-like small GTP-binding protein (smg GDS)
102	EST00248	OG	rho H12/ ARH12
1834	EST01620	OM	AMP deaminase, brain
691	EST00675	OM	Alcohol dehydrogenase
396	EST01443	OM	CDPdiacylglycerol-serine O-phosphatidyltransferase
2192	EST01257	OM	Diacylglycerol kinase, lymphyocyte
1441	EST02477	OM	Diamine acetyltransferase
2289	EST01325	OM	Fatty acid synthase
1020	EST02034	OM	Glutaminase
2326	EST01791	OM	Inositol-1,4,5-trisphosphate 3-kinase
1427	EST02463	OM	Long-chain-fatty-acid-CoA ligase
2226	EST01744	OM	NAD(P)+ transhydrogenase (B-specific)
1566	EST02609	OM	Neutrophil oxidase factor
1681	EST01573	OM	Nucleoside diphosphate kinase

SEQ ID	EST#	Group	Putative Identification
2254	EST01751	OM	Phosphatidylinositol-4,5-bisphosphate phosphodiest
1654	EST01572	OM	Protochlorophyllide reductase
2073	EST01697	OM	Succinate dehydrogenase flavoprotein
2138	EST01715	OM	Succinate dehydrogenase flavoprotein
1771	EST01601	OM	Thiosulfate sulfurtransferase (rhodanese)
2173	EST01724	PI	Lon protease
2297	EST01775	PI	Prohormone cleavage enzyme
9	EST00376	PI	Prolyl endopeptidase
1726	EST01588	PI	XPR2 alkaline extracellular protease
1147	EST02169	PP	Tyrosine kinase
2282	EST01764	RT	Lamin B receptor
189	EST00282	RT	trkB
251	EST00370	SC	Actin, other
2146	EST01218	SC	Actin, other
248	EST00271	SC	Actinin, alpha
891	EST01891	SC	Actinin, alpha
1500	EST02538	SC	Actinin, alpha
132	EST00110	SC	Agrin
1852	EST01625	SC	Agrin
2004	EST01676	SC	Cofilin
650	EST00642	SC	Dilute (myosin heavy chain)
2217	EST01738	SC	Gelation factor ABP-280
77	EST00257	SC	Kinesin
78	EST00258	SC	Kinesin
2245	EST01748	SC	Kinesin
1468	EST02505	SC	Matrin 3
223	EST00368	SC	Microtubule-associated protein 1B
824	EST01865	SC	Microtubule-associated protein 1B
2032	EST01683	SC	Microtubule-associated protein 1B
506	EST01471	SC	Neuraxin
951	EST01960	SC	Spectrin, beta
1371	EST02402	SC	Talin
653	EST01512	SC	Tubulin, alpha
311	EST00270	SC	Tubulin, beta
594	EST01490	SC	Tubulin, beta
757	EST01542	SC	Tubulin, beta
1245	EST02274	SC	Tubulin, beta
1589	EST02634	SC	Tubulin, beta
1701	EST00853	SC	Unc-104
969	EST01982	ST	ADP-ribosylation factor 1
1126	EST02146	ST	Calbindin D28
1910	EST01645	ST	Calmodulin
161	EST00247	ST	MARCKS (myristoylated alanine-rich protein kinase
769	EST00734	ST	MARCKS homolog
1386	EST02418	ST	MARCKS homolog
1931	EST01041	ST	cAMP-regulated phosphoprotein
1413	EST02447	ST	cAMP-specific phosphodiesterase
299	EST00249	ST	smg p25A GDP dissociation inhibitor

SEQ ID	EST#	Group	Putative Identification
2092	EST01700	TP	Anion exchanger homolog AE3
1956	EST01663	TP	Ca2+-transporting ATPase 2
1039	EST02055	TP	Calcium channel
1667	EST00825	TP	Gamma-aminobutyric acid transporter
1412	EST02446	TP	Glutamate-aspartate carrier protein
913	EST01913	TT	Clathrin coat assembly protein AP50 homolog
1035	EST02051	TT	J1 protein
93	EST00287	TT	Processing enhancing protein
38	EST00374	TT	RNA polymerase II 6th subunit (RPO26)
1715	EST01583	TT	Ribosomal protein L18a
1856	EST01627	TT	Ribosomal protein L1a
1974	EST01667	TT	Ribosomal protein L3
301	EST00300	TT	Ribosomal protein L30
22	EST00301	TT	Ribosomal protein S10
2402	EST01826	TT	Ribosomal protein S10
463	EST01459	TT	Ribosomal protein YL10
2121	EST01711	TT	Valine-tRNA ligase
1332	EST02362	TX	GA binding protein, beta subunit
1229	EST02258	TX	KUP protein
1395	EST02429	TX	Nuclear factor 1-like protein (NF1)
187	EST00152	TX	Wilm's tumor-related protein
249	EST00275	TX	Zinc Finger Proteins
413	EST01446	TX	Zinc Finger Proteins
469	EST01460	TX	Zinc Finger Proteins
833	EST01560	TX	Zinc Finger Proteins
1230	EST02259	TX	Zinc finger proteins
1496	EST02534	TX	Zinc finger proteins
2324	EST01352	TX	Zinc Finger Proteins

Group Key: CS: Cell Surface, DC: Developmental Control, EM: Energy Metabolism, KP: Kinases and Phosphatases, OG: Oncogenes, OM: Other Metabolism, PI, Peptidases and Peptidase Inhibitors, RT: Receptors, SC: Structural and Cytoskeletal, ST: Signal Transduction, TP: Transporters, TT: Transcription, Translation, and Subcellular Localization, TX: Transcription Factors.

EXAMPLE 11

cDNA Libraries Generated From Specific Genomic DNA by Exon Expression & Amplification

Exon amplification was used to express potential exons from genomic DNA in a recombinant vector that contains some of the signals necessary for splicing. If an exon is present in the proper orientation in the vector, that exon will be spliced in a mammalian cell and will become part of the mRNA of that cell. The exon splice-product can be purified from other mRNA in the cell by conversion of the mRNA to cDNA and selective amplification of the recombinant splice-product cDNAs. Cosmid DNA from human chromosome 19q13.3 was digested with BamHI or BamHI/Bg1II restriction enzymes. The fragments generated were collected and size specifically cloned into an expression vector (Buckler, et al. **Proc. Nat'l. Acad. Sci. USA,** 88:4005-4009 (1991)). After transfection by electroporation of these constructs into COS cells, RNA transcripts were generated using the SV40 early promoter and a polyadenylation signal derived from SV40 both present in the expression vector. When a fragment of genomic DNA contains an entire exon with flanking intron sequence in the sense orientation, the exon should be retained in the mature poly(A)+ cytoplasmic RNA. Therefore, the mRNA was used as template for cDNA synthesis using reverse transcriptase and vector-priming. Subsequently, the cDNAs were amplified by vector-priming using PCR. A fraction of this first PCR product was reamplified using internal vector-primers containing terminal cloning sites. These products were end-repaired with T4 DNA polymerase, digested with the appropriate restriction enzymes, gel purified and cloned into pBluescript vectors. The constructs were transfected into XL1-Blue competent cells and plated on LB/X-gal/IPTG/ampicillin plates. White colonies were selected and expanded to prepare DNA templates as described in Example 2.

When multiple cosmids or YAC clones were used as the source DNA, a pool of specific expressed exons was obtained as a cDNA library. The EST/cDNAs sequenced from this specific library are disclosed herein as SEQ ID NOS: 2412-2417.

EXAMPLE 12

PCR Amplification from Predicted Exons

Computational analyses can be applied to genomic DNA sequences to predict protein coding regions. The coding region prediction program CRM (E. Uberbacher and R. Mural, **Proc. Natl. Acad. Sci. USA 88**:11261-5 (1991)) finds open reading frames and classifies them according to their probability of being coding regions. These regions are subsequently examined using the GM program (C. Fields and C. Soderlund, **Comp. Applic. Biosci. 6**: 263, 1990), which predicts intron-exon structure. PCR primers are then designed to amplify the predicted exons and used to test human cDNA libraries (for example, fetal brain or placental libraries) for the presence of these putative exons using a PCR assay.

This strategy has been successfully applied in two large scale genomic sequencing projects, the Huntington's locus of human chromosome 4p16.3 (McCombie, et al., submitted) and human chromosome locus 19q13.3 (Martin-Gallardo, et al., submitted). Sequences from eleven predicted exons from chromosome 4 were present in tested cDNA libraries, indicating that this region has at least two and probably three expressed genes. In one case, the method resulted in an amplification product which spanned two predicted exons. (SEQ ID NO: 2411.) When sequenced, this PCR product indicated the presence of the two exons from which the primers were initially chosen, as well as an intervening exon which was also predicted by the CRM program, but not the intervening genomic sequences. In a similar fashion, the presence of the two predicted genes in the chromosome 19

sequence was confirmed by sequencing PCR products. SEQ ID NO 2410, includes a partial exon of one of these genes.

EXAMPLE 13

Complete Sequence of EST Clone Inserts

There are a number of methods known to those with skill in the art of molecular biology, to obtain sequence information from the cDNAs corresponding to the EST sequences. Procedures for these methods are provided in Basic Methods in Molecular Biology (David et al. supra). One way to acquire more information about the cDNA from which an EST was derived is to sequence the remainder of the cDNA clone. The complete sequence of the inserts of four EST clones (representing SEQ ID NOs 188, 189, 223, and 227) was determined using Exonuclease III deletions. Briefly, EST clones were digested with the restriction enzymes SalI and KpnI or PstI and BamHI (for deletions from the Forward primer and Reverse primer ends of the insert, respectively). The KpnI and PstI enzymes leave 3' sticky ends following digestion, which Exonuclease III is unable to bind. This results in unidirectional deletions into the cDNA insert leaving the vector sequence undisturbed. After addition of Exonuclease III to the Forward and Reverse deletion reactions, aliquots of the reaction were removed at defined time intervals and the reaction was stopped to prevent further deletion. S1 nuclease and Klenow DNA polymerase were added to create blunt ended fragments suitable for ligation.

Samples for each time point was purified by electrophoresis through an agarose gel and religated. Two to four representative clones from each time point in each direction were sequenced to give between 200 and 400 base pairs of sequence data. Careful selection of deletion conditions and time points allow a deletion series of approximately 100-200 base pairs difference in length at each consecutive time point. Sequence fragments were reassembled into a redundant contiguous sequence using the INHERIT

software from Applied Biosystems, Inc. (Foster City, CA). In this way, the complete insert from these four cDNA clones was sequenced on both strands to an average redundancy between three and four (each base was sequenced between three and four times, on average). Those complete insert sequences are disclosed herein as SEQ ID 2418, 2419, 2420, and 2421, corresponding to original ESTs with SEQ ID 223, 189, 227, and 188, respectively.

EXAMPLE 14

Determining Reading Frame, Orientation, Coding Regions: ESTs and Complete cDNA Sequences

Once the complete cDNA sequence has been determined in accordance with Example 13, the reading frame, orientation, and coding regions are determined by computer techniques. (The complete coding region is considered to be the largest open reading frame from a methionine to a stop codon.)

Specifically, the CRM program on the GRAIL server is used as explained in Example 9 to determine probable coding regions. This information is supplemented by location of start and stop codons. Where possible, the results of the CRM analysis are validated by comparison of the cDNA sequence to known sequences using database matching, in accordance with Examples 3 and 4. If a match of 50% (or even less) is found in any particular reading frame and orientation, this serves to verify corresponding CRM results. Alternatively, database matches can be used to determine reading frame and orientation without use of the CRM program. Of course, if the cDNA is derived from a directional library, the probable orientation is already known.

EXAMPLE 15

Preparation of PCR Primers and Amplification of DNA

The EST sequences and the corresponding cDNA sequences and genomic sequences may be used, in accordance with the

present invention, to prepare PCR primers for a variety of applications. The PCR primers are preferably at least 15 bases, and more preferably at least 18 bases in length. The procedure of Example 5 is repeated using the desired EST, or using the corresponding cDNA or genomic DNA sequence from Example 13. It is preferred that the primer pairs have approximately the same G/C ratio, so that melting temperatures are approximately the same. When screening cDNA, introns are of no concern; however, when screening genomic DNA, primers should be selected to avoid reading across introns, which usually are too large to amplify. The PCR primers and amplified DNA of this Example find use in the Examples that follow.

EXAMPLE 16

Forensic Matching by DNA Sequencing

In one exemplary method, DNA samples are isolated from forensic specimens of, for example, hair, semen, blood or skin cells by conventional methods. A panel of PCR primers derived from a number of the sequences of Example 1, 2, 11, 12 and/or 13 is then utilized in accordance with Example 12 to obtain DNA of approximately 100-200 bases in length from the forensic specimen. Corresponding sequences are obtained from a suspect. Each of these identification DNAs is then sequenced, and a simple database comparison determines the differences, if any, between the sequences from the suspect and those from the sample. Statistically significant differences between the suspect's DNA sequences and those from the sample conclusively prove a lack of identity. This lack of identity can be proven, for example, with only one sequence. Identity, on the other hand, should be demonstrated with a large number of sequences, all matching. Preferably, a minimum of 50 statistically identical sequences

of 100 bases in length are used to prove identity between the suspect and the sample.

EXAMPLE 17

Positive Identification by DNA Sequencing

The technique outlined in the previous example may also be used on a larger scale to provide a unique fingerprint-type identification of any individual. In this technique, primers are prepared from a large number of sequences from Examples 1, 2, 11, 12 and/or 13. Preferably, 20 to 50 different primers are used. These primers are used to obtain a corresponding number of PCR-generated DNA segments from the individual in question in accordance with Example 15. Each of these DNA segments is sequenced, using the methods set forth in Example 1. The database of sequences generated through this procedure uniquely identifies the individual from whom the sequences were obtained. The same panel of primers may then be used at any later time to absolutely correlate tissue or other biological specimen with that individual.

EXAMPLE 18

Southern Blot Forensic Identification

The procedure of Example 17 is repeated to obtain a panel of from 10 to 2000 amplified sequences from an individual and a specimen. This PCR-generated DNA is then digested with one or a combination of, preferably, four base specific restriction enzymes. Such enzymes are commercially available and known to those of skill in the art. After digestion, the resultant gene fragments are size separated in multiple duplicate wells on an agarose gel and transferred to nitrocellulose using Southern blotting techniques well known to those with skill in the art. For a review of Southern

blotting see Davis et al. (Basic Methods in Molecular Biology, 1986, Elsevier Press. pp 62-65).

A panel of ESTs or complete cDNA sequences from Examples 1, 2, and/or 13, or fragments thereof of at least 15 bases, are radioactively or colorimetrically labeled using end-labeled oligonucleotides derived from the ESTs, nick translated sequences or the like using methods known in the art and hybridized to the Southern blot using techniques known in the art (Davis et al., supra). Preferably, at least 5 to 10 of these labeled probes are used, and more preferably at least about 20 or 30 are used to provide a unique pattern. The resultant bands appearing from the hybridization of a large sample of ESTs will be a unique identifier. Since the restriction enzyme cleavage will be different for every individual, the band pattern on the Southern blot will also be unique. Increasing the number of EST probes will provide a statistically higher level of confidence in the identification since there will be an increased number of sets of bands used for identification.

EXAMPLE 19

Dot Blot Identification Procedure

Another technique for identifying individuals using the sequences disclosed herein utilizes a dot blot hybridization technique.

Genomic DNA is isolated from nuclei of subject to be identified. Oligonucleotide probes of approximately 30 bp in length were synthesized that correspond to sequences from the ESTs. The probes are used to hybridize to the genomic DNA through conditions known to those in the art. The oligonucleotides are end labelled with P^{32} using polynucleotide kinase (Pharmacia). Dot Blots are created by spotting about 50 ng cDNA of at least 10, preferably at least 50 sequences corresponding to a variety of the Sequence ID

NOs provided in Table 7 onto nitrocellulose or the like using a vacuum dot blot manifold (BioRad, Richmond California). The nitrocellulose filter containing the EST clone sequences is baked or UV linked to the filter, prehybridized and hybridized with labeled probe using techniques known in the art (Davis et al. supra). The ^{32}P labeled DNA fragments are sequentially hybridized with successively stringent conditions to detect minimal differences between the 30 bp sequence and the DNA. Tetramethylammonium chloride is useful for identifying clones containing small numbers of nucleotide mismatches (Wood et al., **Proc. Natl. Acad. Sci. USA 82(6)**:1585-1588 (1985) which is hereby incorporated by reference. A unique pattern of dots distinguishes one individual from another individuals.

EXAMPLE 20

Alternative "Fingerprint" Identification Technique

EST sequences and the corresponding complete cDNA sequences can be used to create a unique fingerprint for an individual. Thus pools of EST sequences can be used in forensics, paternity suits or the like to differentiate one individual from another.

Entire EST sequences can be used; similarly oligonucleotides can be prepared from EST sequences. In this example, 20-mer oligonucleotides are prepared from 200 EST sequences using commercially available oligonucleotide services such as Oligos Etc., Wilsonville, OR. Patient cell samples are processed for DNA using techniques well known to those with skill in the art. The nucleic acid is digested with restriction enzymes EcoRI and XbaI. Following digestion, samples are applied to wells for electrophoresis. The procedure, as known in the art, may be modified to accommodate polyacrylamide electrophoresis, however in this example, samples containing 5 ug of DNA are loaded into wells

and separated on 0.8% agarose gels. The gels are transferred using Southern blotting techniques onto nitrocellulose.

10 ng of each of the oligos are pooled and end-labeled with P^{32}. The nitrocellulose is prehybridized with blocking solution and hybridized with the labeled probes. Following hybridization and washing, the nitrocellulose filter is exposed to X-Omat AR X-ray film. The resulting hybridization pattern will be unique for each individual.

It is additionally contemplated within this example that the representative number of EST sequences can be varied for additional accuracy or clarity.

EXAMPLE 21

Identification of genes associated with hereditary diseases

This example illustrates an approach useful for the association of EST sequences with particular phenotypic characteristics. In this example, a particular EST is used as a test probe to associate that EST with a particular phenotypic characteristic.

An EST clone corresponding to EST01643, (SEQ ID NO 1894) maps to a gene rich region of chromosome 6. EST clone HHCMH89, from which EST01643 was derived, was mapped to chromosome 6p21 by Dr. Julie Korenberg of UCLA/Cedar Sinai Hospital using FISH. A search of Mendelian Inheritance in Man (supra) revealed 6p21 to be a very gene rich region containing several known genes and several diseases for which genes have not been identified. The cDNA encoded by EST clone HHCMH89 thus becomes an immediate candidate for each of these genetic diseases.

Cells from patients with these diseases are isolated and expanded in culture. PCR primers from the EST sequences are used to screen genomic DNA and RNA or cDNA from the patients. ESTs that are not amplified in the patients can be positively associated with a particular disease by further analysis.

EXAMPLE 22

Identification of a gene associated with Angelman's disease

Angelman's disease (AD) is characterized by deletions on the long arm of chromosome 15 (15q11q13) (Williams et al. **Am. J. Med. Genet.** 32:339-345 (1989) hereby incorporated by reference). The symptoms of the disease include developmental delay, seizures, inappropriate laughter and ataxic movements. These symptoms suggest that the disorder is a neurologic deficiency. This prophetic example illustrates how ESTs, preferably obtained from a cDNA library from human brain, may be used in identifying the defective gene or genes associated with Angelman's Disease. (The example is based on analogous work with genomic DNA, rather than cDNA and ESTs, in identifying the genetic defect associated with Angelman's Disease.) This example also illustrates how EST sequences may generally be used for identifying gene sequences associated with an inherited disease that is mapped to a chromosome location.

ESTs are screened using techniques described in Example 5 and Example 7 to identify those ESTs that localize to the long arm of chromosome 15 and preferably localize to chromosome 15 bands 15q11q13 from normal patients. ESTs that bind to the long arm of chromosome 15 are hybridized to chromosome 15 from AD patients. These studies are preferrably performed using either fluorescence in situ hybridization or using somatic cell hybrids that contain fragments from the long arm of chromosome 15 from AD patients. Those chromosome 15-specific ESTs that do not map to chromosome 15 from AD patients are useful as markers for Angelman's Disease and can be incorporated into diagnostics for genetic screening. These ESTs are associated with chromosome deletions present in Angelman's disease. Identification of the gene associated with these AD negative ESTs and an analysis of the polypeptides encoded by the genes

from normal patients is essential for providing gene or other therapies for AD patients.

Genetic diseases are not always accompanied by gene deletions. Therefore, it is also important to use the ESTs that bind to bands 15q11q13 from AD patients as tools to identify the polymorphisms present within the disease population. Restriction fragment length polymorphism (RFLP) analysis can be performed on patient cells from AD disease or from somatic cell hybrids created using the long arm of chromosome 15. For a review of RFLP techniques see Donis-Keller et al. (**Cell 51**:319-337 (1987) hereby incorporated by reference). DNA is isolated from the somatic cell lines or from cells from AD patients. The DNA is digested with one or more restriction enzymes according to techniques of Donis-Keller et al. The resulting fragments are separated by gel electrophoresis, denatured, transferred to nitrocellulose and hybridized with the selected radio-labeled ESTs that localize to the region of interest. The autoradiographic pattern is compared both to a number of AD patients and to normal patients. Common patterns of EST hybridization in AD patients that are not present in normal patients indicates that the genes associated with these ESTs are candidate genes affected by AD.

cDNA libraries are prepared from the somatic cell hybrids from AD patients. Libraries are prepared using Lambda Zap II Library Kits (Stratagene, La Jolla, California) or other commercially available library kits. The ESTs of interest are used as probes to identify those bacterial colonies carrying genes corresponding to the EST probes. Positive clones are sequenced and the sequences are compared to homologous gene sequences derived from normal patients.

Alterations, including deletions and substitutions, within gene sequences, associated with bands 15q11q13, are thus positively identified and associated with AD disease. Wagstaff et al. were able to identify deletions and substitutions in sequences encoding the $GABA_A$ receptor

protein subunit from patients with Angelman's disease (**Am. J. Hum. Genet. 49**:330-337, (1991)). It is likely that other genes will additionally be associated with the disease.

EXAMPLE 23

Preparation and Use of Antisense Oligonucleotides

Antisense RNA molecules are known to be useful for regulating translation within the cell. Antisense RNA molecules can be produced from EST sequences or from the corresponding gene sequences. These antisense molecules can be used as diagnostic probes to determine whether or not a particular gene is expressed in a cell. Similarly, the antisense molecules can be used as a therapeutic to regulate gene expression once the EST is associated with a particular disease (see Example 22).

The antisense molecules are obtained from a nucleotide sequence by reversing the orientation of the coding region with regard to the promoter. Thus, the antisense RNA is complementary to the corresponding mRNA. For a review of antisense design see Green et al., **Ann. Rev. Biochem. 55**:569-597 (1986), which is hereby incorporated by reference. The antisense sequences can contain modified sugar phosphate backbones to increase stability and make them less sensitive to RNase activity. Examples of the modifications are described by Rossi et al., **Pharmacol. Ther. 50(2)**:245-254, (1991).

Antisense molecules are introduced into cells that express the gene corresponding to the EST of interest in culture. In a preferred application of this invention, the polypeptide encoded by the gene is first identified, so that the effectiveness of antisense inhibition on translation can be monitored using techniques that include but are not limited to antibody-mediated tests such as RIAs and ELISA, functional assays, or radiolabelling. The antisense molecule is introduced into the cells by diffusion or by transfection

procedures known in the art. The molecules are introduced onto cell samples at a number of different concentrations preferably between $1x10^{-10}$M to $1x10^{-4}$M. Once the minimum concentration that can adequately control translation is identified, the optimized dose is translated into a dosage suitable for use in vivo. For example, an inhibiting concentration in culture of $1x10^{-7}$ translates into a dose of approximately 0.6 mg/kg bodyweight. Levels of oligonucleotide approaching 100 mg/kg bodyweight or higher may be possible after testing the toxicity of the oligonucleotide in laboratory animals.

The antisense can be introduced into the body as a bare or naked oligonucleotide, oligonucleotide encapsulated in lipid, oligonucleotide sequence encapsidated by viral protein, or as oligonucleotide contained in an expression vector such as those described in Example 25. The antisense oligonucleotide is preferably introduced into the vertebrate by injection. It is additionally contemplated that cells from the vertebrate are removed, treated with the antisense oligonucleotide, and reintroduced into the vertebrate. It is further contemplated that the antisense oligonucleotide sequence is incorporated into a ribozyme sequence to enable the antisense to bind and cleave its target. For technical applications of ribozyme and antisense oligonucleotides see Rossi et al.

EXAMPLE 24

Preparation and use of Triple Helix Probes

Triple helix oligonucleotides are used to inhibit transcription from a genome. They are particularly useful for studying alterations in cell activity as it is associated with a particular gene. The EST sequences or complete sequences of the present invention or, more preferably, a portion of those sequences, can be used to inhibit gene

expression in individuals having diseases associated with a particular gene. Similarly, a portion of the EST or corresponding gene sequence can be used to study the effect of inhibiting transcription of a particular gene within a cell. Traditionally, homopurine sequences were considered the most useful. However, homopyrimidine sequences can also inhibit gene expression. Thus, both types of sequences from either the EST or from the gene corresponding to the EST are contemplated within the scope of this invention. Homopyrimidine oligonucleotides bind to the major groove at homopurine:homopyrimidine sequences. As an example, 10-mer to 20-mer homopyrimidine sequences from the ESTs can be used to inhibit expression from homopurine sequences. SEQ ID NOs such as 282, 888, 719, 670, 994, 240, 873 and 761 contain homopyrimidine 15-mers. Moreover the natural (beta) anomers of the oligonucleotide units can be replaced with alpha anomers to render the oligonucleotide more resistant to nucleases. Further, an intercalating agent such as ethidium bromide, or the like, can be attached to the 3' end of the alpha oligonucleotide to stabilize the triple helix. For information on the generation of oligonucleotides suitable for triple helix formation see Griffin et al. (**Science 245**:967-971 (1989), which is hereby incorporated by this reference).

The oligonucleotides may be prepared on an oligonucleotide synthesizer or they may be purchased commercially from a company specializing in custom oligonucleotide synthesis. The sequences are introduced into cells in culture using techniques known in the art that include but are not limited to calcium phosphate precipitation, DEAE-Dextran, electroporation, liposome-mediated transfection or native uptake. Treated cells are monitored for altered cell function. These cell functions are predicted based upon the homologies of the gene, corresponding to the EST from which the oligonucleotide was derived, with known genes sequences that have been associated

with a particular function. The cell functions can also be predicted based on the presence of abnormal physiologies within cells derived from individuals with a particular inherited disease, particularly when the EST is associated with the disease using techniques described in Example 22.

EXAMPLE 25

Gene expression from DNA Sequences Corresponding to ESTs

A gene sequence of the present invention coding for all or part of a human gene product is introduced into an expression vector using conventional technology. (Techniques to transfer cloned sequences into expression vectors that direct protein translation in mammalian, yeast, insect or bacterial expression systems are well known in the art.) Commercially available vectors and expression systems are available from a variety of suppliers including Stratagene (La Jolla, California), Promega (Madison, Wisconsin), and Invitrogen (San Diego, California). If desired, to enhance expression and facilitate proper protein folding, the codon context and codon pairing of the sequence may be optimized for the particular expression organism, as explained by Hatfield, et al., U.S. Patent No. 5,082,767, incorporated herein by this reference.

The following is provided as one exemplary method to generate polypeptide from cloned cDNA sequences. The cDNA from the EST of interest is sequenced to identify the methionine initiation codon for the gene and the poly A sequence. If the cDNA lacks a poly A sequence, this sequence can be added to the construct by, for example, splicing out the Poly A sequence from pSG5 (Stratagene) using BglI and SalI restriction endonuclease enzymes and incorporating it into the mammalian expression vector pXT1 (Stratagene). pXT1 contains the LTRs and a portion of the *gag* gene from Moloney Murine Leukemia Virus. The position of the LTRs in the construct allow efficient stable transfection. The vector

includes the Herpes Simplex Thymidine Kinase promoter and the selectable neomycin gene. The cDNA is obtained by PCR from the bacterial vector using oligonucleotide primers complementary to the cDNA and containing restriction endonuclease sequences for Pst I incorporated into the 5'primer and BglII at the 5' end of the corresponding cDNA 3' primer, taking care to ensure that the cDNA is positioned inframe with the poly A sequence. The purified fragment obtained from the resulting PCR reaction is digested with PstI, blunt ended with an exonuclease, digested with Bgl II, purified and ligated to pXT1, now containing a poly A sequence and digested BglII.

The ligated product is transfected into mouse NIH 3T3 cells using Lipofectin (Life Technologies, Inc., Grand Island, New York) under conditions outlined in the product specification. Positive transfectants are selected after growing the transfected cells in 600ug/ml G418 (Sigma, St. Louis, Missouri). The protein is preferrably released into the supernatant. However if the protein has membrane binding domains, the protein may additionally be retained within the cell or expression may be restricted to the cell surface.

Since it may be necessary to purify and locate the transfected product, synthetic 15-mer peptides synthesized from the predicted cDNA sequence are injected into mice to generate antibody to the polypeptide encoded by the cDNA.

If antibody production is not possible, the cDNA sequence is additionally incorporated into eukaryotic expression vectors and expressed as a chimeric with, for example, β-globin. Antibody to β-globin is used to purify the chimeric. Corresponding protease cleavage sites engineered between the β-globin gene and the cDNA are then used to separate the two polypeptide fragments from one another after translation. One useful expression vector for generating β-globin chimerics is pSG5 (Stratagene). This vector encodes rabbit β-globin. Intron II of the rabbit β-globin gene facilitates splicing of the expressed transcript,

and the polyadenylation signal incorporated into the construct increases the level of expression. These techniques as described are well known to those skilled in the art of molecular biology. Standard methods are published in methods texts such as Davis et al. and many of the methods are available from the technical assistance representatives from Stratagene, Life Technologies, Inc., or Promega. Polypeptide may additionally be produced from either construct using in vitro translation systems such as In vitro Express™ Translation Kit (Stratagene).

Example 26

Production of an Antibody to a Human Protein

Substantially pure protein or polypeptide is isolated from the transfected or transformed cells as described in Example 25. Concentration of protein in the final preparation is adjusted, for example, by concentration on an Amicon filter device, to the level of a few micrograms/ml. Monoclonal or polyclonal antibody to the protein can then be prepared as follows:

A. Monoclonal Antibody Production by Hybridoma Fusion

Monoclonal antibody to epitopes of any of the peptides identified and isolated as described can be prepared from murine hybridomas according to the classical method of Kohler, G. and Milstein, C., **Nature 256**:495 (1975) or derivative methods thereof. Briefly, a mouse is repetitively inoculated with a few micrograms of the selected protein over a period of a few weeks. The mouse is then sacrificed, and the antibody producing cells of the spleen isolated. The spleen cells are fused by means of polyethylene glycol with mouse myeloma cells, and the excess unfused cells destroyed by growth of the system on selective media comprising aminopterin (HAT media). The successfully fused cells are diluted and aliquots of the dilution placed in wells of a

microtiter plate where growth of the culture is continued. Antibody-producing clones are identified by detection of antibody in the supernatant fluid of the wells by immunoassay procedures, such as Elisa, as originally described by Engvall, E., **Meth. Enzymol. 70:419 (1980)**, and derivative methods thereof. Selected positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Davis, L. et al. **Basic Methods in Molecular Biology** Elsevier, New York. Section 21-2.

B. Polyclonal Antibody Production by Immunization

Polyclonal antiserum containing antibodies to heterogenous epitopes of a single protein can be prepared by immunizing suitable animals with the expressed protein described above, which can be unmodified or modified to enhance immunogenicity. Effective polyclonal antibody production is affected by many factors related both to the antigen and the host species. For example, small molecules tend to be less immunogenic than other and may require the use of carriers and adjuvant. Also, host animals vary in response to site of inoculations and dose, with both inadequate or excessive doses of antigen resulting in low titer antisera. Small doses (ng level) of antigen administered at multiple intradermal sites appears to be most reliable. An effective immunization protocol for rabbits can be found in Vaitukaitis, J. et al. **J. Clin. Endocrinol. Metab. 33:988-991 (1971)**.

Booster injections can be given at regular intervals, and antiserum harvested when antibody titer thereof, as determined semi-quantitatively, for example, by double immunodiffusion in agar against known concentrations of the antigen, begins to fall. See, for example, Ouchterlony, O. et al., Chap. 19 in: **Handbook of Experimental Immunology** D. Wier (ed) Blackwell (1973). Plateau concentration of antibody is usually in the range of 0.1 to 0.2 mg/ml of serum (about 12 μM). Affinity of the antisera for the antigen is

determined by preparing competitive binding curves, as described, for example, by Fisher, D., Chap. 42 in: **Manual of Clinical Immunology**, 2d Ed. (Rose and Friedman, eds.) Amer. Soc. For Microbiol., Washington, D.C. (1980).

Antibody preparations prepared according to either protocol are useful in quantitative immunoassays which determine concentrations of antigen-bearing substances in biological samples; they are also used semi-quantitatively or qualitatively to identify the presence of antigen in a biological sample.

EXAMPLE 27

Identification of Tissue Types or Cell Species by Means of Labeled Tissue Specific Antibodies

Identification of specific tissues is accomplished by the visualization of tissue specific antigens by means of antibody preparations according to Example 26 which are conjugated, directly or indirectly to a detectable marker. Selected labeled antibody species bind to their specific antigen binding partner in tissue sections, cell suspensions, or in extracts of soluble proteins from a tissue sample to provide a pattern for qualitative or semi-qualitative interpretation.

Antisera for these procedures must have a potency exceeding that of the native preparation, and for that reason, antibodies are concentrated to a mg/ml level by isolation of the gamma globulin fraction, for example, by ion-exchange chromatography or by ammonium sulfate fractionation. Also, to provide the most specific antisera, unwanted antibodies, for example to common proteins, must be removed from the gamma globulin fraction, for example by means of insoluble immunoabsorbents, before the antibodies are labeled with the marker. Either monoclonal or heterologous antisera is suitable for either procedure.

A. Immunohistochemical Techniques

Purified, high-titer antibodies, prepared as described above, are conjugated to a detectable marker, as described, for example, by Fudenberg, H., Chap. 26 in: **Basic & Clinical Immunology**, 3rd Ed. Lange, Los Altos, California (1980) or Rose, N. et al., Chap. 12 in: **Methods in Immunodiagnosis**, 2d Ed. John Wiley & Sons, New York (1980).

A fluorescent marker, either fluorescein or rhodamine, is preferred, but antibodies can also be labeled with an enzyme that supports a color producing reaction with a substrate, such as horseradish peroxidase. Markers can be added to tissue-bound antibody in a second step, as described below. Alternatively, the specific antitissue antibodies can be labeled with ferritin or other electron dense particles, and localization of the ferritin coupled antigen-antibody complexes achieved by means of an electron microscope. In yet another approach, the antibodies are radiolabeled, with, for example ^{125}I, and detected by overlaying the antibody treated preparation with photographic emulsion.

Preparations to carry out the procedures can comprise monoclonal or polyclonal antibodies to a single gene copy or protein, identified as specific to a tissue type, for example, brain tissue, or antibody preparations to several antigenically distinct tissue specific antigens can be used in panels, independently or in mixtures, as required.

Tissue sections and cell suspensions are prepared for immunohistochemical examination according to common histological techniques. Multiple cryostat sections (about 4 μm, unfixed) of the unknown tissue and known control, are mounted and each slide covered with different dilutions of the antibody preparation. Sections of known and unknown tissues should also be treated with preparations to provide a positive control, a negative control, for example, pre-immune sera, and a control for non-specific staining, for example, buffer.

Treated sections are incubated in a humid chamber for 30 min at room temperature, rinsed, then washed in buffer for 30-45 min. Excess fluid is blotted away, and the marker developed.

If the tissue specific antibody was not labeled in the first incubation, it can be labeled at this time in a second antibody-antibody reaction, for example, by adding fluorescein- or enzyme-conjugated antibody against the immunoglobulin class of the antiserum-producing species, for example, fluorescein labeled antibody to mouse IgG. Such labeled sera are commercially available.

The antigen found in the tissues by the above procedure can be quantified by measuring the intensity of color or fluorescence on the tissue section, and calibrating that signal using appropriate standards.

B. Identification of Tissue Specific Soluble Proteins

The visualization of tissue specific proteins and identification of unknown tissues from that procedure is carried out using the labeled antibody reagents and detection strategy as described for immunohistochemistry; however the sample is prepared according to an electrophoretic technique to distribute the proteins extracted from the tissue in an orderly array on the basis of molecular weight for detection.

A tissue sample is homogenized using a Virtis apparatus; cell suspensions are disrupted by Dounce homogenization or osmotic lysis, using detergents in either case as required to disrupt cell membranes, as is the practice in the art. Insoluble cell components such as nuclei, microsomes, and membrane fragments are removed by ultracentrifugation, and the soluble protein-containing fraction concentrated if necessary and reserved for analysis.

A sample of the soluble protein solution is resolved into individual protein species by conventional SDS polyacrylamide electrophoresis as described, for example, by Davis, L. et al., Section 19-2 in: **Basic Methods in Molecular Biology** (P. Leder, ed), Elsevier, New York (1986), using a

range of amounts of polyacrylamide in a set of gels to resolve the entire molecular weight range of proteins to be detected in the sample. A size marker is run in parallel for purposes of estimating molecular weights of the constituent proteins. Sample size for analysis is a convenient volume of from 5-50 μl, and containing from about 1 to 100 μg protein. An aliquot of each of the resolved proteins is transferred by blotting to a nitrocellulose filter paper, a process that maintains the pattern of resolution. Multiple copies are prepared. The procedure, known as Western Blot Analysis, is well described in Davis, L. et al., (above) Section 19-3. One set of nitrocellulose blots is stained with Coomassie Blue dye to visualize the entire set of proteins for comparison with the antibody bound proteins. The remaining nitrocellulose filters are then incubated with a solution of one or more specific antisera to tissue specific proteins prepared as described in Example 26. In this procedure, as in procedure A above, appropriate positive and negative sample and reagent controls are run.

In either procedure A or B, a detectable label can be attached to the primary tissue antigen-primary antibody complex according to various strategies and permutations thereof. In a straightforward approach, the primary specific antibody can be labeled; alternatively, the unlabeled complex can be bound by a labeled secondary anti-IgG antibody. In other approaches, either the primary or secondary antibody is conjugated to a biotin molecule, which can, in a subsequent step, bind an avidin conjugated marker. According to yet another strategy, enzyme labeled or radioactive protein A, which has the property of binding to any IgG, is bound in a final step to either the primary or secondary antibody.

The visualization of tissue specific antigen binding at levels above those seen in control tissues to one or more tissue specific antibodies, prepared from the gene sequences identified from EST sequences, can identify tissues of unknown origin, for example, forensic samples, or

differentiated tumor tissue that has metastasized to foreign bodily sites.

The entire contents of all references cited above are hereby incorporated by reference.

While the present invention has been described in some detail for purposes of clarity and understanding, one skilled in the art will appreciate that various changes in form and detail can be made without departing from the true scope of the invention.

VII. Correlation of EST and Clone Identifiers

The EST sequences of the present invention are identified herein by SEQ ID NO, and are identified in the GenBank database by a different number, are identified in the inventors' lab (and upcoming publications) by EST number, and clones have been submitted to the American Type Culture Collection (Rockville, Maryland USA) under clone names. Table 12 cross references those different numbers for the ESTs from cDNA, SEQ ID NOS 1-2409.

Certain Sequence ID NOS are excluded from some claims based on their homology to known non-human sequences (See Table 2).

Table 12. SEQ ID NO Cross References

SEQ ID	EST#	GB#	Clone
1	EST00007	M61959	HFBA01
2	EST00009	M61953	HFBA05
3	EST00010	M61961	HFBA07
4	EST00011	M61962	HFBA08
5	EST00012	M61963	HFBA10
6	EST00013	M61964	HFBA11
7	EST00014	M61965	HFBA26
8	EST00234	M62172	HFBA20
9	EST00376	M61966	HFBA23
10	EST00016	M61967	HFBA36
11	EST00018	M61968	HFBA51
12	EST00274	M62212	HFBA65
13	EST00055	M62194	HFBA66
14	EST00019	M61969	HFBA69
15	EST00020	M61970	HFBA71
16	EST00021	M61971	HFBA77
17	EST00022	M62299	HFBA84
18	EST00373	M61973	HFBA86
19	EST00023	M61974	HFBA87
20	EST00024	M61975	HFBA89
21	EST00025	M62239	HFBA90
22	EST00301	M61976	HHCA08
23	EST00026	M61977	HHCA05
24	EST00027	M61978	HHCA08
25	EST00028	M62245	HHCA02
26	EST00310	M61979	HHCA08
27	EST00029	M61980	HHCA09
28	EST00030	M61981	HHCA104
29	EST00031	M61982	HHCA10
30	EST00032	M61983	HHCA11
31	EST00033	M61171	HHCA11
32	EST00233	M61984	HHCA13
33	EST00034	M61985	HHCA14
34	EST00035	M61986	HHCA18
35	EST00036	M61987	HHCA21
36	EST00037	M61988	HHCA23
37	EST00038	M62300	HHCA53
38	EST00374	M61989	HHCA53
39	EST00039	M61990	HHCA54
40	EST00040	M61991	HHCA54
41	EST00041	M61992	HHCA55
42	EST00042	M62297	HHCA57
43	EST00371	M62290	HHCA68
44	EST00364	M61994	HHCA69
45	EST00044	M61995	HHCA71
46	EST00046	M62229	HHCA73
47	EST00291	M61996	HHCA79
48	EST00047	M61997	HHCB04
49	EST00048	M61998	HHCB08
50	EST00049	M61999	HHCB16
51	EST00052	M62000	HHCB35
52	EST00054	M62001	HHCB51
53	EST00055	M62002	HHCB52
54	EST00056	M62003	HHCB53
55	EST00057	M62004	HHCB55
56	EST00058	M62005	HHCB96
57	EST00059	M62006	HHCC04
58	EST00061	M62007	HHCC05
59	EST00062	M62008	HHCC07
60	EST00290	M62228	HHCC09
61	EST00064	M62009	HHCC10
62	EST00065	M62010	HHCC13
63	EST00066	M62011	HHCC18
64	EST00067	M62280	HHCC21
65	EST00351	M62012	HHCC22
66	EST00068	M62287	HHCC23
67	EST00340	M62014	HHCC27
68	EST00070	M62015	HHCC29
69	EST00071	M62016	HHCC31
70	EST00072	M62019	HHCC37
71	EST00073	M62018	HHCC40
72	EST00074	M62019	HHCC42
73	EST00075	M62196	HHCC53
74	EST00257	M62197	HHCC53
75	EST00258	M62020	HHCC64
76	EST00076	M62021	HHCC67
77	EST00077	M62248	HHCC70
78	EST00315	M62022	HHCC72
79	EST00078	M62023	HHCC74
80	EST00079	M62024	HHCC76
81	EST00080	M62025	HHCC77
82	EST00081	M62026	HHCC78
83	EST00082	M62027	HHCC80
84	EST00083	M62028	HHD001
85	EST00084	M62029	HHD002
86	EST00085	M62240	HHD003
87	EST00302	M62030	HHD005
88	EST00086	M62031	HHD006
89	EST00087	M62225	HHD008
90	EST00287	M62281	HHD010
91	EST00088	M62032	HHD012
92	EST00089	M62033	HHD014
93	EST00289	M62227	HHD017
94	EST00290	M62226	HHD021
95	EST00316	M62249	HHD030
96	EST00090	M62034	HHE004
97	EST00091	M62035	HHE005
98	EST00248	M62187	HHE006
99	EST00317	M62250	HHE007
100	EST00354	M62282	HHE008
101	EST00345	M62291	HHE010
102	EST00092	M62036	HHE012
103	EST00093	M62037	HHE013
104	EST00094	M62038	HHE015
105	EST00095	M62039	HHE017
106	EST00096	M62040	HHE020
107	EST00281	M62218	HHG001
108	EST00318	M62251	HHG003
109	EST00097	M62041	HHG007
110	EST00098	M62042	HHG010
111	EST00099	M62043	HHG014
112	EST00100	M62044	HHG018
113	EST00319	M62252	HHG025
114	EST00101	M62045	HHG029
115	EST00102	M62046	HHG031
116	EST00103	M62047	HHG033
117	EST00104	M62048	HHG036
118	EST00105	M62049	HHG037
119	EST00106	M62050	HHG038
120	EST00107	M62051	HHG040
121	EST00108	M62052	HHG041
122	EST00109	M62053	HHG044
123	EST00320	M62253	HHG051
124	EST00321	M62054	HHG053
125	EST00110	M62055	HHG055
126	EST00111	M62186	HHG057
127	EST00375	M62056	HHG060
128	EST00252	M62191	HHG063
129	EST00321	M62254	HHG065
130	EST00355	M62283	HHG066
131	EST00322	M62255	HHG067
132	EST00110	M62057	HHG068
133	EST00111	M62186	HHG070
134	EST00375	M62058	HHG072
135	EST00112	M62056	HHG085
136	EST00113	M62057	HHH001
137	EST00272	M62210	HHH009
138	EST00114	M62058	HHH011
139	EST00116	M62059	HHH015
140	EST00117	M62060	HHH036
141	EST00118	M62061	HHH041
142	EST00323	M62256	HHH049
143	EST00119	M62063	HHH050
144	EST00120	M62062	HHH051
145	EST00121	M62064	HHH052
146	EST00122	M62065	HHH053
147	EST00292	M62230	HHH055
148	EST00236	M62174	HHH062
149	EST00123	M62066	HHH063
150	EST00124	M62067	HHH064
151	EST00125	M62068	HHH066
152	EST00126	M62069	HHH068
153	EST00127	M62070	HHH071
154	EST00128	M62071	HHH072
155	EST00129	M62072	HHH073
156	EST00130	M62073	HHH076
157	EST00131	M62074	HHH077
158	EST00132	M62075	HHH079
159	EST00325	M62257	HHH080
160	EST00326	M62258	HHH081
161	EST00047	M62185	HHH082
162	EST00133	M62076	HHH083
163	EST00134	M62077	HHH084
164	EST00135	M62078	HHH087
165	EST00136	M62079	HHH091
166	EST00137	M62080	HHH096
167	EST00138	M62081	HHI105
168	EST00140	M62082	HHI106
169	EST00141	M62083	HHI109
170	EST00295	M62233	HHI110
171	EST00327	M62259	HHI117
172	EST00142	M62084	HHI126
173	EST00143	M62085	HHI127
174	EST00296	M62234	HHI133
175	EST00144	M62086	HHI136
176	EST00328	M62260	HHI142
177	EST00294	M62232	HHI147
178	EST00145	M62087	HHI154
179	EST00299	M62237	HHI155
180	EST00146	M62088	
181	EST00147		

SEQ ID	EST#	GB#	Clone
182	EST00329	M62261	HHC159
183	EST00148	M62089	HHC161
184	EST00149	M62090	HHC162
185	EST00150	M62091	HHC173
186	EST00151	M62092	HHC175
187	EST00152	M62093	HHC179
188	EST00256	M62195	HHC184
189	EST00282	M62219	HHC185
190	EST00153	M62094	HHC186
191	EST00154	M62095	HHC188
192	EST00155	M62096	HHC190
193	EST00156	M62097	HHC192
194	EST00157	M62098	HHC193
195	EST00158	M62099	HHCJ05
196	EST00159	M62100	HHCJ07
197	EST00160	M62101	HHCJ09
198	EST00161	M62102	HHCJ13
199	EST00277	M62214	HHCJ15
200	EST00162	M62103	HHCJ17
201	EST00163	M62104	HHCJ29
202	EST00298	M62236	HHCJ30
203	EST00164	M62105	HHCJ34
204	EST00235	M62173	HHCJ35
205	EST00165	M62106	HHCJ36
206	EST00166	M62107	HHCJ37
207	EST00167	M62108	HHCJ42
208	EST00250	M62189	HHCJ43
209	EST00331	M62262	HHCJ47
210	EST00168	M62109	HHCJ48
211	EST00332	M62263	HHCJ50
212	EST00169	M62110	HHCJ51
213	EST00170	M62111	HHCJ59
214	EST00171	M62112	HHCJ60
215	EST00172	M62113	HHCJ61
216	EST00173	M62114	HHCJ67
217	EST00174	M62116	HHCJ73
218	EST00175	M62117	HHCJ74
219	EST00176	M62298	HHCJ78
220	EST00372	M62286	HHCJ79
221	EST00359	M62118	HHCJ80
222	EST00177	M62294	HHCJ84
223	EST00368	M62284	HHCJ85
224	EST00356	M62119	HHCJ86
225	EST00178	M62264	HHCJ91
226	EST00333	M62120	HHCJ92
227	EST00259	M62198	HHCJ96
228	EST00179	M62121	HHCSA26
229	EST00180	M62122	HHCSA28
230	EST00181	M62265	HHCSA29
231	EST00334	M62123	HHCSA30
232	EST00182	M62124	HHCSA31
233	EST00183	M62125	HHCSA49
234	EST00184	M62126	HHCSA53
235	EST00185	M62127	HHCSA59
236	EST00186	M62128	HHCSA64
237	EST00187	M62129	HHCSA70
238	EST00188	M62130	HHCSA75
239	EST00189	M62266	HHCSA77
240	EST00335	M62131	HHCSA79
241	EST00191	M62132	HHCSA81
242	EST00192	M62133	HHCSA82
243	EST00193	M62134	HHCSA85
244	EST00194	M62135	
245	EST00347	M62275	
246	EST00196	M62217	
247	EST00279	M62209	
248	EST00271		

SEQ ID	EST#	GB#	Clone
249	EST00275	M62213	HHCSA86
250	EST00197	M62136	HHCSA87
251	EST00370	M62296	HHCSA92
252	EST00198	M62137	HHCSA95
253	EST00199	M62138	HHCSB05
254	EST00200	M62139	HHCSB07
255	EST00201	M62140	HHCSB08
256	EST00345	M62267	HHCSB09
257	EST00337	M62268	HHCSB12
258	EST00346	M62276	HHCSB13
259	EST00202	M62141	HHCSB19
260	EST00357	M62285	HHCSB20
261	EST00338	M62269	HHCSB21
262	EST00339	M62270	HHCSB22
263	EST00203	M62142	HHCSB23
264	EST00204	M62143	HHCSB25
265	EST00205	M62144	HHCSB27
266	EST00206	M62145	HHCSB29
267	EST00297	M62235	HHCSB30
268	EST00369	M62295	HHCSB31
269	EST00293	M62231	HHCSB32
270	EST00207	M62146	HHCSB34
271	EST00283	M62220	HHCSB37
272	EST00340	M62271	HHCSB43
273	EST00208	M62147	HHCSB44
274	EST00268	M62206	HHCSB45
275	EST00209	M62148	HHCSB46
276	EST00211	M62149	HHCSB48
277	EST00212	M62150	HHCSB50
278	EST00342	M62272	HHCSB53
279	EST00213	M62151	HHCSB55
280	EST00343	M62273	HHCSB56
281	EST00214	M62152	HHCSB57
282	EST00344	M62274	HHCSB58
283	EST00215	M62153	HHCSB62
284	EST00216	M62154	HHCSB68
285	EST00286	M62224	HHCSB69
286	EST00217	M62155	HHCSB70
287	EST00218	M62156	HHCSB73
288	EST00219	M62157	HHCSB77
289	EST00220	M62158	HHCSB78
290	EST00221	M62159	HHCSB79
291	EST00222	M62160	HHCSB81
292	EST00223	M62161	HHCSB83
293	EST00224	M62162	HHCSB84
294	EST00225	M62163	HHCSB91
295	EST00226	M62164	HHCSB92
296	EST00228	M62166	HTCA08
297	EST00231	M62168	HTCA27
298	EST00230	M62169	HTCA27
299	EST00249	M62188	HTCA36
300	EST00232	M62170	HTCB03
301	EST00300	M62238	HTCB05
302	EST00303	M62241	HTCB22
303	EST00348	M62277	HFBA45
304	EST00307	M62242	HFBA78
305	EST00308	M62243	HHCA05
306	EST00309	M62244	HHCA06
307	EST00312	M62246	HHCA57
308	EST00314	M62247	HHCA63
309	EST00174	M62115	HHCJ67
310	EST00377		HHCG05
311	EST00270	M62208	HFBA18
312	EST00276	M62215	HHCJ13
313	EST00008	M61960	HFBA04
315	EST00378	M78230	HEFBA01
316			

SEQ ID	EST#	GB#	Clone
317	EST00379	M78231	HEFBA01
318	EST00380	M78232	HEFBA04
319	EST00381	M78233	HEFBA04
320	EST00382	M78234	HEFBA07
321	EST00383	M78235	HEFBA07
322	EST00384	M78236	HEFBA09
323	EST00385	M78237	HEFBA10
324	EST01827	M85319	HEFBA11
325	EST00386	M78238	HEFBA13
326	EST00387	M78239	HEFBA13
327	EST00388	M78240	HEFBA18
328	EST00389	M78241	HEFBA18
329	EST00390	M78242	HEFBA21
330	EST00391	M78243	HEFBA23
331	EST00392	M78244	HEFBB04
332	EST00393	M78245	HEFBB04
333	EST00394	M78246	HEFBB09
334	EST00395	M78247	HEFBB21
335	EST00396	M78248	HEFBB24
336	EST00397	M78249	HEFBB30
337	EST00398	M78250	HEFBC02
338	EST00399	M78251	HEFBC06
339	EST00400	M78252	HEFBC11
340	EST00402	M78254	HEFBC17
341	EST00403	M78255	HEFBC21
342	EST00404	M78256	HEFBD03
343	EST01428	M79273	HEFBD03
344	EST00405	M78257	HEFBD05
345	EST00406	M78258	HEFBD09
346	EST01828	M85320	HEFBE18
347	EST01829	M85321	HEFBF14
348	EST01830	M85322	HFBA86
349	EST01831	M85323	HFBA90
350	EST00407	M78259	HFBBA01
351	EST00408	M78260	HFBBA02
352	EST00409	M78261	HFBBA03
353	EST00410	M78262	HFBBA04
354	EST01433	M79278	HFBBA05
355	EST00411	M78263	HFBBA06
356	EST00412	M78264	HFBBA07
357	EST00413	M78265	HFBBA08
358	EST00414	M78266	HFBBA09
359	EST00415	M78267	HFBBA10
360	EST00416	M78268	HFBBA11
361	EST00417	M78269	HFBBA14
362	EST00418	M78270	HFBBA15
363	EST00419	M78271	HFBBA16
364	EST00420	M78272	HFBBA17
365	EST01434	M85318	HFBBA18
366	EST00421	M78273	HFBBA20
367	EST00422	M78274	HFBBA21
368	EST00423	M78275	HFBBA22
369	EST00424	M78276	HFBBA23
370	EST00425	M78277	HFBBA24
371	EST00426	M78278	HFBBA25
372	EST00427	M78279	HFBBA26
373	EST01832	M85324	HFBBA28

SEQ ID	EST#	GB#	Clone
507	EST00530	M78382	HFBCB13
508	EST00531	M78383	HFBCB14
509	EST01472	M77888	HFBCB15
510	EST00532	M78384	HFBCB16
511	EST00533	M78385	HFBCB19
512	EST00534	M78386	HFBCB20
513	EST00535	M78387	HFBCB21
514	EST00536	M78388	HFBCB22
515	EST00537	M78389	HFBCB23
516	EST00538	M78390	HFBCB24
517	EST00539	M78391	HFBCB26
518	EST00540	M78392	HFBCB28
519	EST00541	M78393	HFBCB31
520	EST00542	M78394	HFBCB33
521	EST01474	M77890	HFBCB34
522	EST00543	M78395	HFBCB35
523	EST01858	M85329	HFBCB36
524	EST00544	M78396	HFBCB38
525	EST00545	M78397	HFBCB40
526	EST00546	M78398	HFBCB42
527	EST00547	M78399	HFBCB43
528	EST00548	M78400	HFBCB46
529	EST00549	M78401	HFBCB48
530	EST01477	M77893	HFBCB49
531	EST00550	M78402	HFBCB51
532	EST00551	M78403	HFBCB52
533	EST00552	M78404	HFBCB53
534	EST01478	M77894	HFBCB54
535	EST00553	M78405	HFBCB55
536	EST01479	M77895	HFBCB56
537	EST00554	M78406	HFBCB57
538	EST00555	M78407	HFBCB59
539	EST00556	M78408	HFBCB60
540	EST00557	M78409	HFBCB61
541	EST00558	M78410	HFBCB62
542	EST01480	M77896	HFBCB63
543	EST00559	M78411	HFBCB64
544	EST00560	M78412	HFBCB65
545	EST01481	M77897	HFBCB66
546	EST01859	M85330	HFBCB67
547	EST00561	M78413	HFBCB69
548	EST00562	M78414	HFBCB71
549	EST00563	M78415	HFBCB73
550	EST00564	M78416	HFBCB74
551	EST01482	M77898	HFBCB75
552	EST00565	M78417	HFBCB76
553	EST00566	M78418	HFBCB77
554	EST00567	M78419	HFBCB79
555	EST01483	M77899	HFBCB80
556	EST00568	M78420	HFBCB81
557	EST00569	M78421	HFBCB82
558	EST01484	M77900	HFBCB83
559	EST00570	M78422	HFBCB84
560	EST00571	M78423	HFBCB85
561	EST01485	M77901	HFBCB87
562	EST00572	M78424	HFBCB89
563	EST00573	M78425	HFBCB91

SEQ ID	EST#	GB#	Clone
440	EST01454	M77870	HFBCA22
441	EST00481	M78333	HFBCA24
442	EST01456	M77872	HFBCA26
443	EST00482	M78334	HFBCA27
444	EST00483	M78335	HFBCA28
445	EST00484	M78336	HFBCA30
446	EST00485	M78337	HFBCA31
447	EST00486	M78338	HFBCA32
448	EST00487	M78339	HFBCA33
449	EST00488	M78340	HFBCA34
450	EST00489	M78341	HFBCA35
451	EST00490	M78342	HFBCA36
452	EST00491	M78343	HFBCA37
453	EST00492	M78344	HFBCA38
454	EST00493	M78345	HFBCA39
455	EST00494	M78346	HFBCA40
456	EST01835	M85326	HFBCA41
457	EST00495	M78347	HFBCA42
458	EST00496	M78348	HFBCA43
459	EST00497	M78349	HFBCA44
460	EST01457	M77873	HFBCA45
461	EST01836	M85327	HFBCA47
462	EST00498	M78350	HFBCA48
463	EST01459	M77875	HFBCA49
464	EST00499	M78351	HFBCA50
465	EST00500	M78352	HFBCA51
466	EST00501	M78353	HFBCA52
467	EST00502	M78354	HFBCA54
468	EST00503	M78355	HFBCA56
469	EST01460	M77876	HFBCA58
470	EST00504	M78356	HFBCA59
471	EST00505	M78357	HFBCA60
472	EST00506	M78358	HFBCA63
473	EST00507	M78359	HFBCA64
474	EST00508	M78360	HFBCA65
475	EST00509	M78361	HFBCA70
476	EST01463	M77879	HFBCA71
477	EST00510	M78362	HFBCA72
478	EST00511	M78363	HFBCA73
479	EST01464	M77880	HFBCA74
480	EST00512	M78364	HFBCA76
481	EST01465	M77881	HFBCA77
482	EST00513	M78365	HFBCA79
483	EST00514	M78366	HFBCA80
484	EST01466	M77882	HFBCA82
485	EST00515	M78367	HFBCA83
486	EST00516	M78368	HFBCA84
487	EST00517	M78369	HFBCA85
488	EST00518	M78370	HFBCA86
489	EST00519	M78371	HFBCA87
490	EST00520	M78372	HFBCA88
491	EST00521	M78373	HFBCA90
492	EST00522	M78374	HFBCA91
493	EST00523	M78375	HFBCA92
494	EST00524	M78376	HFBCA94
495	EST00525	M78377	HFBCA95
496	EST00526	M78378	HFBCA96
497	EST01467	M77883	HFBCB01
498	EST01468	M77884	HFBCB02
499	EST00527	M78379	HFBCB03
500	EST02715		HFBCB05
501	EST01469	M77885	HFBCB06
502	EST00528	M78380	HFBCB08
503	EST00529	M78381	HFBCB09
504	EST01837	M85328	HFBCB10
505	EST01471	M77887	HFBCB12
506			

SEQ ID	EST#	GB#	Clone
374	EST00428	M78280	HFBBA30
375	EST00429	M78281	HFBBA32
376	EST01436	M77852	HFBBA33
377	EST00430	M78282	HFBBA34
378	EST00431	M78283	HFBBA37
379	EST00432	M78284	HFBBA38
380	EST01439	M77855	HFBBA40
381	EST00433	M78285	HFBBA41
382	EST00434	M78286	HFBBA42
383	EST00435	M78287	HFBBA43
384	EST01440	M77856	HFBBA44
385	EST00436	M78288	HFBBA45
386	EST00437	M78289	HFBBA47
387	EST00438	M78290	HFBBA48
388	EST00439	M78291	HFBBA49
389	EST00440	M78292	HFBBA51
390	EST01442	M77858	HFBBA52
391	EST00441	M78293	HFBBA53
392	EST00442	M78294	HFBBA54
393	EST00443	M78295	HFBBA56
394	EST00444	M78296	HFBBA57
395	EST00445	M78297	HFBBA58
396	EST01443	M77859	HFBBA59
397	EST00446	M78298	HFBBA60
398	EST00447	M78299	HFBBA61
399	EST00448	M78300	HFBBA62
400	EST00449	M78301	HFBBA63
401	EST00450	M78302	HFBBA64
402	EST00451	M78303	HFBBA65
403	EST00452	M78304	HFBBA66
404	EST00453	M78305	HFBBA67
405	EST00454	M78306	HFBBA68
406	EST00455	M78307	HFBBA69
407	EST00456	M78308	HFBBA72
408	EST00457	M78309	HFBBA73
409	EST01444	M77860	HFBBA74
410	EST00458	M78310	HFBBA76
411	EST00459	M78311	HFBBA77
412	EST01445	M77861	HFBBA80
413	EST01446	M77862	HFBBA81
414	EST00460	M78312	HFBBA82
415	EST00461	M78313	HFBBA83
416	EST00462	M78314	HFBBA84
417	EST00463	M78315	HFBBA85
418	EST00464	M78316	HFBBA87
419	EST00465	M78317	HFBBA89
420	EST00466	M78318	HFBBA90
421	EST00467	M78319	HFBBA91
422	EST01447	M77863	HFBBA92
423	EST00468	M78320	HFBBA93
424	EST01448	M77864	HFBBA96
425	EST00469	M78321	HFBCA01
426	EST00470	M78322	HFBCA02
427	EST01449	M77865	HFBCA04
428	EST01451	M77867	HFBCA06
429	EST00471	M78323	HFBCA07
430	EST00472	M78324	HFBCA08
431	EST00473	M78325	HFBCA09
432	EST01452	M77868	HFBCA11
433	EST00474	M78326	HFBCA12
434	EST00475	M78327	HFBCA13
435	EST00476	M78328	HFBCA15
436	EST00477	M78329	HFBCA16
437	EST00478	M78330	HFBCA17
438	EST00479	M78331	HFBCA18
439	EST00480	M78332	HFBCA21

SEQ ID	EST#	GB#	Clone
564	EST00574	M78426	HFBCB92
565	EST00575	M78427	HFBCB93
566	EST00576	M78428	HFBCB94
567	EST00577	M78429	HFBCB96
568	EST00578	M78430	HFBCC01
569	EST00579	M78431	HFBCC02
570	EST00580	M78432	HFBCC03
571	EST00581	M78433	HFBCC04
572	EST00582	M78434	HFBCC06
573	EST00583	M78435	HFBCC09
574	EST00584	M78436	HFBCC11
575	EST00585	M78437	HFBCC13
576	EST00586	M78438	HFBCC14
577	EST00587	M78439	HFBCC15
578	EST00588	M78440	HFBCC16
579	EST00589	M78441	HFBCC17
580	EST00590	M78442	HFBCC18
581	EST00591	M78443	HFBCC20
582	EST00592	M78444	HFBCC21
583	EST00593	M78445	HFBCC22
584	EST00594	M78446	HFBCC23
585	EST00595	M78447	HFBCC25
586	EST00596	M78448	HFBCC26
587	EST01488	M77904	HFBCC27
588	EST00597	M78449	HFBCC28
589	EST00598	M78450	HFBCC29
590	EST00599	M78451	HFBCC30
591	EST01489	M77905	HFBCC31
592	EST00600	M78452	HFBCC32
593	EST00601	M78453	HFBCC33
594	EST01490	M77906	HFBCC34
595	EST01840	M85331	HFBCC35
596	EST00602	M78454	HFBCC36
597	EST00603	M78455	HFBCC37
598	EST00604	M78456	HFBCC38
599	EST00605	M78457	HFBCC39
600	EST01492	M77908	HFBCC40
601	EST01493	M77909	HFBCC41
602	EST00606	M78458	HFBCC42
603	EST01494	M77910	HFBCC43
604	EST00607	M78459	HFBCC44
605	EST00608	M78460	HFBCC46
606	EST00609	M78461	HFBCC47
607	EST00610	M78462	HFBCC50
608	EST00611	M78463	HFBCC51
609	EST01496	M77912	HFBCC52
610	EST00612	M78464	HFBCC53
611	EST00613	M78465	HFBCC54
612	EST00614	M78466	HFBCC55
613	EST01842	M85332	HFBCC56
614	EST00615	M78467	HFBCC57
615	EST00616	M78468	HFBCC63
616	EST01497	M77913	HFBCC65
617	EST00617	M78469	HFBCC66
618	EST01498	M77914	HFBCC67
619	EST00619	M78471	HFBCC69
620	EST01499	M77915	HFBCC70
621	EST00620	M78472	HFBCC71
622	EST01843	M85333	HFBCC72
623	EST00621	M78473	HFBCC73
624	EST01500	M77916	HFBCC75
625	EST01844	M85334	HFBCC77
626	EST00622	M78474	HFBCC80
627	EST00623	M78475	HFBCC81
628	EST01503	M77919	HFBCC82
629	EST00624	M78476	HFBCC84

SEQ ID	EST#	GB#	Clone
630	EST01505	M77921	HFBCC85
631	EST00625	M78477	HFBCC89
632	EST00626	M78478	HFBCC90
633	EST00627	M78479	HFBCC91
634	EST00628	M78480	HFBCC92
635	EST00629	M78481	HFBCC93
636	EST01507	M77923	HFBCC94
637	EST00630	M78482	HFBCC95
638	EST00631	M78483	HFBCD03
639	EST01432	M78484	HFBCD04
640	EST01509	M77925	HFBCD06
641	EST00633	M78485	HFBCD07
642	EST00634	M78486	HFBCD08
643	EST00635	M78487	HFBCD09
644	EST00636	M78488	HFBCD10
645	EST00637	M78489	HFBCD11
646	EST00638	M78490	HFBCD12
647	EST00639	M78491	HFBCD16
648	EST00640	M78492	HFBCD17
649	EST00641	M78493	HFBCD18
650	EST00642	M78494	HFBCD19
651	EST00643	M78495	HFBCD20
652	EST01510	M77926	HFBCD21
653	EST01512	M77928	HFBCD26
654	EST00644	M78496	HFBCD27
655	EST00645	M78497	HFBCD28
656	EST01513	M77929	HFBCD29
657	EST00646	M78498	HFBCD31
658	EST00647	M78499	HFBCD33
659	EST00648	M78500	HFBCD34
660	EST00649	M78501	HFBCD35
661	EST00650	M78502	HFBCD36
662	EST00651	M78503	HFBCD37
663	EST00652	M78504	HFBCD38
664	EST00653	M78505	HFBCD39
665	EST00654	M78506	HFBCD40
666	EST01514	M77930	HFBCD41
667	EST00655	M78507	HFBCD42
668	EST00656	M78508	HFBCD43
669	EST00657	M78509	HFBCD44
670	EST00658	M78510	HFBCD45
671	EST00659	M78511	HFBCD46
672	EST00660	M78512	HFBCD47
673	EST01515	M77931	HFBCD48
674	EST01516	M77932	HFBCD49
675	EST00661	M78513	HFBCD51
676	EST00662	M78514	HFBCD54
677	EST00663	M78515	HFBCD57
678	EST01517	M77933	HFBCD61
679	EST01518	M77934	HFBCD62
680	EST00664	M78516	HFBCD63
681	EST00665	M78517	HFBCD64
682	EST00666	M78518	HFBCD65
683	EST00667	M78519	HFBCD68
684	EST00668	M78520	HFBCD69
685	EST00669	M78521	HFBCD70
686	EST00670	M78522	HFBCD72
687	EST00671	M78523	HFBCD74
688	EST00672	M78524	HFBCD76
689	EST00673	M78525	HFBCD79
690	EST00674	M78526	HFBCD80
691	EST00675	M78527	HFBCD81
692	EST00676	M78528	HFBCD82
693	EST00677	M78529	HFBCD83
694	EST00678	M78530	HFBCD84
695	EST00679	M78531	HFBCD86

SEQ ID	EST#	GB#	Clone
696	EST01521	M77937	HFBCD87
697	EST00680	M78532	HFBCD88
698	EST00681	M78533	HFBCD89
699	EST00682	M78534	HFBCD90
700	EST00683	M78535	HFBCD91
701	EST01522	M77938	HFBCD92
702	EST00684	M78536	HFBCD94
703	EST00685	M78537	HFBCD95
704	EST00686	M78538	HFBCD96
705	EST00687	M78539	HFBCE01
706	EST00688	M78540	HFBCE02
707	EST01847	M85335	HFBCE03
708	EST00689	M78541	HFBCE05
709	EST00690	M78542	HFBCE06
710	EST00691	M78543	HFBCE07
711	EST00692	M78544	HFBCE08
712	EST00693	M78545	HFBCE11
713	EST00694	M78546	HFBCE12
714	EST00695	M78547	HFBCE13
715	EST01523	M77939	HFBCE14
716	EST01524	M77940	HFBCE15
717	EST01525	M77941	HFBCE16
718	EST00696	M78548	HFBCE17
719	EST01526	M77942	HFBCE18
720	EST00697	M78549	HFBCE19
721	EST01527	M77943	HFBCE21
722	EST01528	M77944	HFBCE22
723	EST00698	M78550	HFBCE23
724	EST01529	M77951	HFBCE24
725	EST00699	M78551	HFBCE26
726	EST00700	M78552	HFBCE27
727	EST00701	M78553	HFBCE28
728	EST00702	M78554	HFBCE29
729	EST00703	M78555	HFBCE30
730	EST00704	M78556	HFBCE31
731	EST00705	M78557	HFBCE32
732	EST00706	M78558	HFBCE34
733	EST00707	M78559	HFBCE36
734	EST00708	M78560	HFBCE37
735	EST00709	M78561	HFBCE38
736	EST01532	M77948	HFBCE41
737	EST00710	M78562	HFBCE42
738	EST00711	M78563	HFBCE46
739	EST01534	M77950	HFBCE47
740	EST01535	M77951	HFBCE48
741	EST00712	M78564	HFBCE49
742	EST00713	M78565	HFBCE50
743	EST00714	M78566	HFBCE52
744	EST01537	M77953	HFBCE54
745	EST00715	M78567	HFBCE55
746	EST00716	M78568	HFBCE56
747	EST00717	M78569	HFBCE58
748	EST01850	M85337	HFBCE59
749	EST00719	M78571	HFBCE61
750	EST01539	M77955	HFBCE63
751	EST01540	M77956	HFBCE64
752	EST00720	M78572	HFBCE65

SEQ ID	EST#	GB#	Clone
885	EST01885	M85371	HFBCH12
886	EST01886	M85372	HFBCH14
887	EST01887	M85373	HFBCH15
888	EST01888	M85374	HFBCH16
889	EST01889	M85375	HFBCH17
890	EST01890	M85376	HFBCH18
891	EST01891	M85377	HFBCH20
892	EST01892	M85378	HFBCH21
893	EST01893	M85379	HFBCH22
894	EST01894	M85380	HFBCH23
895	EST01895	M85381	HFBCH24
896	EST01896	M85382	HFBCH31
897	EST01897	M85383	HFBCH33
898	EST01898	M85384	HFBCH34
899	EST01899	M85385	HFBCH35
900	EST01900	M85386	HFBCH36
901	EST01901	M85387	HFBCH37
902	EST01902	M85388	HFBCH38
903	EST01903	M85389	HFBCH39
904	EST01904	M85390	HFBCH42
905	EST01905	M85391	HFBCH43
906	EST01906	M85392	HFBCH45
907	EST01907	M85393	HFBCH46
908	EST01908	M85394	HFBCH50
909	EST01909	M85395	HFBCH56
910	EST01910	M85396	HFBCH57
911	EST01911	M85397	HFBCH58
912	EST01912	M85398	HFBCH60
913	EST01913	M85399	HFBCH61
914	EST01914	M85400	HFBCH62
915	EST01915	M85401	HFBCH63
916	EST01916	M85402	HFBCH65
917	EST01917	M85403	HFBCH66
918	EST01918	M85404	HFBCH68
919	EST01919	M85405	HFBCH70
920	EST01920	M85406	HFBCH71
921	EST01921	M85407	HFBCH72
922	EST01922	M85408	HFBCH73
923	EST01923	M85409	HFBCH74
924	EST01924	M85410	HFBCH76
925	EST01925	M85411	HFBCH77
926	EST01926	M85412	HFBCH78
927	EST01927	M85413	HFBCH81
928	EST01929	M85414	HFBCH82
929	EST01930	M85415	HFBCH84
930	EST01931	M85416	HFBCH85
931	EST01932	M85417	HFBCH86
932	EST01933	M85418	HFBCH87
933	EST01934	M85419	HFBCH90
934	EST01935	M85420	HFBCH92
935	EST01936	M85421	HFBCH93
936	EST01937	M85422	HFBCH94
937	EST01938	M85423	HFBCH95
938	EST01939	M85424	HFBCH96
939	EST01940	M85425	HFBC102
940	EST01941	M85426	HFBC105
941	EST01943	M85427	HFBC106
	EST01944	M85428	HFBC108
	EST01945	M85429	

SEQ ID	EST#	GB#	Clone
819	EST01859	M85345	HFBCG01
820	EST01860	M85346	HFBCG02
821	EST01862	M85348	HFBCG09
822	EST01863	M85349	HFBCG10
823	EST01864	M85350	HFBCG11
824	EST01865	M85351	HFBCG12
825	EST01866	M85352	HFBCG13
826	EST01867	M85353	HFBCG15
827	EST01868	M77974	HFBCG17
828	EST00767	M78619	HFBCG19
829	EST00768	M78620	HFBCG20
830	EST01539	M77975	HFBCG21
831	EST00769	M78621	HFBCG22
832	EST00770	M78622	HFBCG23
833	EST01540	M77976	HFBCG24
834	EST00771	M78623	HFBCG25
835	EST00772	M78624	HFBCG26
836	EST00773	M78625	HFBCG27
837	EST01541	M77977	HFBCG29
838	EST00774	M78626	HFBCG30
839	EST01542	M77978	HFBCG31
840	EST00775	M78627	HFBCG32
841	EST00776	M78628	HFBCG33
842	EST01543	M77979	HFBCG34
843	EST01544	M77980	HFBCG35
844	EST01565	M77981	HFBCG37
845	EST00777	M78629	HFBCG38
846	EST00778	M78630	HFBCG40
847	EST00779	M78631	HFBCG43
848	EST01566	M77982	HFBCG44
849	EST01567	M77983	HFBCG45
850	EST00780	M78632	HFBCG47
851	EST00781	M78633	HFBCG49
852	EST00782	M78634	HFBCG51
853	EST00783	M78635	HFBCG53
854	EST00784	M78636	HFBCG57
855	EST00785	M78637	HFBCG61
856	EST01568	M77984	HFBCG62
857	EST01869	M85354	HFBCG69
858	EST01870	M85355	HFBCG72
859	EST00786	M85356	HFBCG73
860	EST01871	M78638	HFBCG74
861	EST01872	M85357	HFBCG76
862	EST01873	M85358	HFBCG77
863	EST00787	M85359	HFBCG78
864	EST01569	M77985	HFBCG79
865	EST01874	M85360	HFBCG80
866	EST01875	M85361	HFBCG81
867	EST01876	M85362	HFBCG83
868	EST00788	M78640	HFBCG84
869	EST00789	M78641	HFBCG85
870	EST00790	M78642	HFBCG88
871	EST00791	M78643	HFBCG89
872	EST00792	M78644	HFBCG90
873	EST00793	M78645	HFBCG92
874	EST00794	M78646	HFBCG93
875	EST00795	M78647	HFBCG94
876	EST00796	M85363	HFBCG96
877	EST01877	M85364	HFBCH01
878	EST01878	M85365	HFBCH02
879	EST01879	M85366	HFBCH03
880	EST01880	M85367	HFBCH05
881	EST01881	M85368	HFBCH06
882	EST01882	M85369	HFBCH07
883	EST01883	M85370	HFBCH08
884	EST01884		HFBCH10

SEQ ID	EST#	GB#	Clone
723	EST00721	M78573	HFBCE66
724	EST00722	M78574	HFBCE69
725	EST00723	M78575	HFBCE72
726	EST01541	M77957	HFBCE73
727	EST01542	M77958	HFBCE74
728	EST00724	M78576	HFBCE77
729	EST00725	M78577	HFBCE78
730	EST00726	M78578	HFBCE80
731	EST01544	M77960	HFBCE82
732	EST00727	M78579	HFBCE83
733	EST00728	M78580	HFBCE84
734	EST00729	M78581	HFBCE85
735	EST00730	M78582	HFBCE86
736	EST00731	M78583	HFBCE87
737	EST00732	M78584	HFBCE88
738	EST00733	M78585	HFBCE90
739	EST00734	M78586	HFBCE91
740	EST00735	M78587	HFBCE93
741	EST01546	M77962	HFBCE94
742	EST00736	M78588	HFBCE95
743	EST01547	M77963	HFBCE96
744	EST01548	M77964	HFBCF01
745	EST00737	M78589	HFBCF03
746	EST00738	M78590	HFBCF07
747	EST00739	M78591	HFBCF09
748	EST00740	M78592	HFBCF10
749	EST00741	M78593	HFBCF11
750	EST01549	M77965	HFBCF13
751	EST01550	M77966	HFBCF14
752	EST01551	M77967	HFBCF16
753	EST01552	M77968	HFBCF23
754	EST01853	M85338	HFBCF41
755	EST00742	M77969	HFBCF42
756	EST00743	M78594	HFBCF43
757	EST00744	M78595	HFBCF44
758	EST00745	M78596	HFBCF45
759	EST01554	M77970	HFBCF46
760	EST00746	M78597	HFBCF47
761	EST00747	M78598	HFBCF48
762	EST00748	M78599	HFBCF49
763	EST01555	M77971	HFBCF50
764	EST00749	M78600	HFBCF51
765	EST00750	M78601	HFBCF52
766	EST00751	M78602	HFBCF53
767	EST01853	M85339	HFBCF54
768	EST00752	M78603	HFBCF56
769	EST00753	M78604	HFBCF57
770	EST00754	M78605	HFBCF58
771	EST00755	M78606	HFBCF60
772	EST00756	M78607	HFBCF61
773	EST00757	M78608	HFBCF63
774	EST00758	M78609	HFBCF68
775	EST00759	M78610	HFBCF73
776	EST00760	M78611	HFBCF74
777	EST00761	M78612	HFBCF75
778	EST00762	M78613	HFBCF79
779	EST00763	M78614	HFBCF81
780	EST00764	M78615	HFBCF84
781	EST01854	M85340	HFBCF85
782	EST00765	M78616	HFBCF86
783	EST00766	M78617	HFBCF87
784	EST01855	M78618	HFBCF89
785	EST01856	M85341	HFBCF90
786	EST01857	M85342	HFBCF91
787	EST01858	M85343	HFBCF93
788		M85344	HFBCF94

SEQ ID	EST#	GB#	Clone
942	EST01947	M85431	HFBCJ11
943	EST01948	M85432	HFBCJ12
944	EST01949	M85433	HFBCJ13
945	EST01950	M85434	HFBCJ14
946	EST01953	M85437	HFBCJ17
947	EST01954	M85438	HFBCJ18
948	EST01957	M85441	HFBCJ23
949	EST01958	M85442	HFBCJ24
950	EST01959	M85443	HFBCJ27
951	EST01960	M85444	HFBCJ28
952	EST01961	M85445	HFBCJ29
953	EST01962	M85446	HFBCJ31
954	EST01963	M85447	HFBCJ32
955	EST01966	M85450	HFBCJ36
956	EST01968	M85452	HFBCJ40
957	EST01969	M85453	HFBCJ41
958	EST01970	M85454	HFBCJ42
959	EST01972	M85456	HFBCJ44
960	EST01973	M85457	HFBCJ45
961	EST01974	M85458	HFBCJ46
962	EST01975	M85459	HFBCJ47
963	EST01976	M85460	HFBCJ48
964	EST01977	M85461	HFBCJ50
965	EST01978	M85462	HFBCJ51
966	EST01979	M85463	HFBCJ52
967	EST01980	M85464	HFBCJ53
968	EST01981	M85465	HFBCJ54
969	EST01982	M85466	HFBCJ55
970	EST01983	M85467	HFBCJ56
971	EST01985	M85469	HFBCJ59
972	EST01986	M85470	HFBCJ60
973	EST01987	M85471	HFBCJ61
974	EST01988	M85472	HFBCJ62
975	EST01989	M85473	HFBCJ63
976	EST01990	M85474	HFBCJ64
977	EST01991	M85475	HFBCJ65
978	EST01992	M85476	HFBCJ67
979	EST01993	M85477	HFBCJ68
980	EST01994	M85478	HFBCJ69
981	EST01995	M85479	HFBCJ70
982	EST01996	M85480	HFBCJ75
983	EST01997	M85481	HFBCJ76
984	EST01998	M85482	HFBCJ77
985	EST01999	M85483	HFBCJ78
986	EST02000	M85484	HFBCJ80
987	EST02001	M85485	HFBCJ81
988	EST02002	M85486	HFBCJ82
989	EST02003	M85487	HFBCJ83
990	EST02004	M85488	HFBCJ84
991	EST02005	M85489	HFBCJ85
992	EST02006	M85490	HFBCJ86
993	EST02007	M85491	HFBCJ87
994	EST02008	M85492	HFBCJ89
995	EST02009	M85493	HFBCJ90
996	EST02010	M85494	HFBCJ91
997	EST02011	M85495	HFBCJ92
998	EST02012	M85496	HFBCJ93
999	EST02013	M85497	HFBCJ95
1000	EST02014	M85498	HFBCJ01
1001	EST02015	M85499	HFBCJ03
1002	EST02016	M85500	HFBCJ04
1003	EST02017	M85501	HFBCJ05
1004	EST02018	M85502	HFBCJ06
1005	EST02019	M85503	HFBCJ09
1006	EST02020	M85504	HFBCJ10
1007	EST02021	M85505	HFBCJ11

SEQ ID	EST#	GB#	Clone
1008	EST02022	M85506	HFBCJ12
1009	EST02023	M85507	HFBCJ17
1010	EST02024	M85508	HFBCJ18
1011	EST02025	M85509	HFBCJ20
1012	EST02026	M85510	HFBCJ34
1013	EST02027	M85511	HFBCJ35
[illegible]	EST02028	M85512	HFBCJ39
[illegible]	EST02029	M85513	HFBCJ40
[illegible]	EST02030	M85514	HFBCJ41
[illegible]	EST02031	M85515	HFBCJ42
1018	EST02032	M85516	HFBCJ49
1019	EST02033	M85517	HFBCJ50
1020	EST02034	M85518	HFBCJ51
1021	EST02035	M85519	HFBCJ52
1022	EST02036	M85520	HFBCJ54
1023	EST02037	M85521	HFBCJ55
1024	EST02038	M85522	HFBCJ56
1025	EST02040	M85524	HFBCJ59
1026	EST02041	M85525	HFBCJ60
1027	EST02042	M85526	HFBCJ61
1028	EST02043	M85527	HFBCJ62
1029	EST02044	M85528	HFBCJ63
1030	EST02045	M85529	HFBCJ64
1031	EST02046	M85530	HFBCJ65
1032	EST02048	M85532	HFBCJ67
1033	EST02049	M85533	HFBCJ68
1034	EST02050	M85534	HFBCJ69
1035	EST02051	M85535	HFBCJ70
1036	EST02052	M85536	HFBCJ72
1037	EST02053	M85537	HFBCJ73
1038	EST02054	M85538	HFBCJ74
1039	EST02055	M85539	HFBCJ75
1040	EST02056	M85540	HFBCJ76
1041	EST02057	M85541	HFBCJ77
1042	EST02058	M85542	HFBCJ78
1043	EST02059	M85543	HFBCJ79
1044	EST02060	M85544	HFBCJ80
1045	EST02061	M85545	HFBCJ85
1046	EST02062	M85546	HFBCJ86
1047	EST02063	M85547	HFBCJ87
1048	EST02064	M85548	HFBCJ90
1049	EST02065	M85549	HFBCJ91
1050	EST02066	M85550	HFBCJ92
1051	EST02067	M85551	HFBCJ94
1052	EST02068	M85552	HFBCJ95
1053	EST02069	M85553	HFBCK01
1054	EST02070	M85554	HFBCK02
1055	EST02071	M85555	HFBCK03
1056	EST02072	M85556	HFBCK04
1057	EST02073	M85557	HFBCK05
1058	EST02074	M85558	HFBCK06
1059	EST02075	M85559	HFBCK07
1060	EST02076	M85560	HFBCK08
1061	EST02078	M85562	HFBCK11
1062	EST02079	M85563	HFBCK12
1063	EST02081	M85565	HFBCK16
1064	EST02082	M85566	HFBCK17
1065	EST02083	M85567	HFBCK18
1066	EST02084	M85568	HFBCK19
1067	EST02085	M85569	HFBCK20
1068	EST02086	M85570	HFBCK21
1069	EST02087	M85571	HFBCK22
1070	EST02088	M85572	HFBCK23
1071	EST02089	M85573	HFBCK25
1072	EST02090	M85574	HFBCK26
1073	EST02091	M85575	HFBCK28

SEQ ID	EST#	GB#	Clone
1074	EST02092	M85576	HFBCK29
1075	EST02093	M85577	HFBCK30
1076	EST02094	M85578	HFBCK31
1077	EST02096	M85580	HFBCK36
1078	EST02097	M85581	HFBCK37
1079	EST02098	M85582	HFBCK38
1080	EST02099	M85583	HFBCK39
1081	EST02100	M85584	HFBCK40
1082	EST02101	M85585	HFBCK42
1083	EST02102	M85586	HFBCK43
1084	EST02103	M85587	HFBCK44
1085	EST02104	M85588	HFBCK45
1086	EST02105	M85589	HFBCK48
1087	EST02106	M85590	HFBCK49
1088	EST02107	M85591	HFBCK54
1089	EST02108	M85592	HFBCK55
1090	EST02109	M85593	HFBCK56
1091	EST02110	M85594	HFBCK57
1092	EST02111	M85595	HFBCK58
1093	EST02112	M85596	HFBCK59
1094	EST02113	M85597	HFBCK60
1095	EST02114	M85598	HFBCK61
1096	EST02115	M85599	HFBCK62
1097	EST02116	M85600	HFBCK63
1098	EST02117	M85601	HFBCK65
1099	EST02118	M85602	HFBCK68
1100	EST02119	M85603	HFBCK69
1101	EST02120	M85604	HFBCK70
1102	EST02121	M85605	HFBCK71
1103	EST02122	M85606	HFBCK72
1104	EST02123	M85607	HFBCK74
1105	EST02124	M85608	HFBCK75
1106	EST02125	M85609	HFBCK76
1107	EST02126	M85610	HFBCK77
1108	EST02127	M85611	HFBCK81
1109	EST02128	M85612	HFBCK83
1110	EST02129	M85613	HFBCK84
1111	EST02131	M85614	HFBCK86
1112	EST02132	M85615	HFBCK87
1113	EST02133	M85616	HFBCK89
1114	EST02134	M85617	HFBCK90
1115	EST02135	M85618	HFBCK91
1116	EST02136	M85619	HFBCK92
1117	EST02137	M85620	HFBCK93
1118	EST02138	M85621	HFBCK94
1119	EST02139	M85622	HFBCK95
1120	EST02140	M85623	HFBCK96
1121	EST02141	M85624	HFBCL01
1122	EST02142	M85625	HFBCL02
1123	EST02143	M85626	HFBCL05
1124	EST02144	M85627	HFBCL07
1125	EST02145	M85628	HFBCL08
1126	EST02146	M85629	HFBCL09
1127	EST02147	M85630	HFBCL10
1128	EST02148	M85631	HFBCL11
1129	EST02149	M85632	HFBCL12
1130	EST02150	M85633	HFBCL13

SEQ ID	EST#	GB#	Clone
1262	EST02281	M85760	HFBCN24
1263	EST02282	M85761	HFBCN25
1264	EST02283	M85762	HFBCN26
1265	EST02284	M85763	HFBCN27
1266	EST02285	M85764	HFBCN36
1267	EST02286	M85765	HFBCN37
1268	EST02287	M85766	HFBCN39
1269	EST02288	M85767	HFBCN40
1270	EST02289	M85768	HFBCN42
1271	EST02290	M85769	HFBCN46
1272	EST02291	M85770	HFBCN47
1273	EST02292	M85771	HFBCN48
1274	EST02293	M85772	HFBCN49
1275	EST02294	M85773	HFBCN50
1276	EST02295	M85774	HFBCN51
1277	EST02296	M85775	HFBCN52
1278	EST02297	M85776	HFBCN53
1279	EST02298	M85777	HFBCN54
1280	EST02299	M85778	HFBCN55
1281	EST02300	M85779	HFBCN56
1282	EST02301	M85780	HFBCN57
1283	EST02302	M85781	HFBCN58
1284	EST02303	M85782	HFBCN59
1285	EST02304	M85783	HFBCN60
1286	EST02305	M85784	HFBCN61
1287	EST02306	M85785	HFBCN62
1288	EST02307	M85786	HFBCN64
1289	EST02308	M85787	HFBCN65
1290	EST02309	M85788	HFBCN66
1291	EST02310	M85789	HFBCN67
1292	EST02311	M85790	HFBCN68
1293	EST02312	M85791	HFBCN69
1294	EST02313	M85792	HFBCN71
1295	EST02314	M85793	HFBCN75
1296	EST02315	M85794	HFBCN77
1297	EST02316	M85795	HFBCN82
1298	EST02317	M85796	HFBCN83
1299	EST02318	M85797	HFBCN84
1300	EST02319	M85798	HFBCN85
1301	EST02320	M85799	HFBCN87
1302	EST02321	M85800	HFBCN88
1303	EST02322	M85801	HFBCN89
1304	EST02323	M85802	HFBCN90
1305	EST02324	M85803	HFBCN91
1306	EST02325	M85804	HFBCN92
1307	EST02326	M85805	HFBCN93
1308	EST02327	M85806	HFBCN94
1309	EST02328	M85807	HFBCN96
1310	EST02329	M85808	HFBD023
1311	EST02330	M85809	HFBD024
1312	EST02331	M85810	HFBD025
1313	EST02332	M85811	HFBD026
1314	EST02333	M85812	HFBD027
1315	EST02334	M85813	HFBD028
1316	EST02335	M85814	HFBD029
1317	EST02336	M85815	HFBD030
1318	EST02337	M85816	HFBD031
1319	EST02338	M85817	HFBD033
1320	EST02339	M85818	HFBD034
1321	EST02340	M85819	HFBD035
1322	EST02341	M85820	HFBD036
1323	EST02342	M85821	HFBD037
1324	EST02343	M85310	

SEQ ID	EST#	GB#	Clone
1167	EST02222	M85704	HFBCN30
1168	EST02223	M85705	HFBCN31
1169	EST02224	M85706	HFBCN32
1170	EST02226	M85707	HFBCN34
1171	EST02228	M85709	HFBCN37
1172	EST02229	M85710	HFBCN38
1173	EST02230	M85711	HFBCN39
1174	EST02232	M85713	HFBCN41
1175	EST02233	M85714	HFBCN42
1176	EST02234	M85715	HFBCN43
1177	EST02235	M85716	HFBCN44
1178	EST02236	M85717	HFBCN45
1179	EST02237	M85718	HFBCN46
1180	EST02238	M85719	HFBCN48
1181	EST02239	M85720	HFBCN49
1182	EST02240	M85721	HFBCN50
1183	EST02241	M85722	HFBCN53
1184	EST02242	M85723	HFBCN54
1185	EST02244	M85725	HFBCN60
1186	EST02245	M85726	HFBCN61
1187	EST02246	M85727	HFBCN64
1188	EST02248	M85728	HFBCN67
1189	EST02249	M85729	HFBCN70
1190	EST02250	M85730	HFBCN72
1191	EST02251	M85731	HFBCN73
1192	EST02252	M85732	HFBCN74
1193	EST02253	M85733	HFBCN75
1194	EST02254	M85734	HFBCN76
1195	EST02255	M85735	HFBCN77
1196	EST02256	M85736	HFBCN78
1197	EST02257	M85737	HFBCN79
1198	EST02258	M85738	HFBCN80
1199	EST02259	M85739	HFBCN81
1200	EST02260	M85740	HFBCN82
1201	EST02261	M85741	HFBCN83
1202	EST02262	M85742	HFBCN84
1203	EST02263	M85743	HFBCN86
1204	EST02264	M85744	HFBCN87
1205	EST02265	M85745	HFBCN88
1206	EST02266	M85746	HFBCN89
1207	EST02267	M85747	HFBCN90
1208	EST02268	M85748	HFBCN91
1209	EST02269	M85749	HFBCN93
1210	EST02270	M85750	HFBCN96
1211	EST02271	M85751	HFBCN14
1212	EST02272	M85752	HFBCN15
1213	EST02273	M85753	HFBCN16
1214	EST02274	M85754	HFBCN17
1215	EST02275	M85755	HFBCN18
1216	EST02276	M85508	HFBCN19
1217	EST02277	M85756	HFBCN20
1218	EST02278	M85509	HFBCN21
1219	EST02279	M85757	HFBCN22
1220	EST02280	M85758	HFBCN23
1221		M85759	

SEQ ID	EST#	GB#	Clone
1131	EST02151	M85634	HFBCL14
1132	EST02152	M85635	HFBCL18
1133	EST02153	M85636	HFBCL20
1134	EST02154	M85637	HFBCL22
1135	EST02155	M85638	HFBCL23
1136	EST02156	M85639	HFBCL24
1137	EST02157	M85640	HFBCL25
1138	EST02159	M85642	HFBCL28
1139	EST02161	M85644	HFBCL30
1140	EST02162	M85645	HFBCL31
1141	EST02163	M85646	HFBCL32
1142	EST02164	M85647	HFBCL34
1143	EST02165	M85648	HFBCL35
1144	EST02166	M85649	HFBCL36
1145	EST02167	M85650	HFBCL38
1146	EST02168	M85651	HFBCL39
1147	EST02169	M85652	HFBCL41
1148	EST02170	M85653	HFBCL46
1149	EST02171	M85654	HFBCL47
1150	EST02172	M85655	HFBCL48
1151	EST02173	M85656	HFBCL50
1152	EST02174	M85657	HFBCL51
1153	EST02175	M85658	HFBCL53
1154	EST02176	M85659	HFBCL54
1155	EST02177	M85660	HFBCL58
1156	EST02178	M85661	HFBCL59
1157	EST02180	M85662	HFBCL64
1158	EST02181	M85663	HFBCL68
1159	EST02182	M85664	HFBCL71
1160	EST02183	M85665	HFBCL72
1161	EST02184	M85666	HFBCL73
1162	EST02185	M85667	HFBCL75
1163	EST02187	M85669	HFBCL77
1164	EST02188	M85670	HFBCL79
1165	EST02189	M85671	HFBCL82
1166	EST02190	M85672	HFBCL83
1167	EST02191	M85673	HFBCL84
1168	EST02193	M85675	HFBCL86
1169	EST02194	M85676	HFBCL87
1170	EST02195	M85677	HFBCL88
1171	EST02196	M85678	HFBCL91
1172	EST02197	M85679	HFBCL92
1173	EST02198	M85680	HFBCL94
1174	EST02199	M85681	HFBCL95
1175	EST02200	M85682	HFBCL96
1176	EST02201	M85683	HFBCM01
1177	EST02202	M85684	HFBCM03
1178	EST02203	M85685	HFBCM05
1179	EST02204	M85686	HFBCM06
1180	EST02205	M85687	HFBCM08
1181	EST02206	M85688	HFBCM09
1182	EST02207	M85689	HFBCM11
1183	EST02208	M85690	HFBCM12
1184	EST02209	M85691	HFBCM14
1185	EST02210	M85692	HFBCM15
1186	EST02211	M85693	HFBCM16
1187	EST02212	M85694	HFBCM17
1188	EST02213	M85695	HFBCM19
1189	EST02214	M85696	HFBCM20
1190	EST02215	M85697	HFBCM21
1191	EST02216	M85698	HFBCM22
1192	EST02217	M85699	HFBCM23
1193	EST02218	M85700	HFBCM25
1194	EST02219	M85701	HFBCM27
1195	EST02220	M85702	HFBCM28
1196	EST02221	M85703	HFBCM29

SEQ ID	EST#	GB#	Clone
1314	EST02344	M85822	HFBC038
1315	EST02345	M85311	HFBC038
1316	EST02346	M85823	HFBC039
1317	EST02347	M85312	HFBC039
1318	EST02348	M85824	HFBC041
1319	EST02349	M85825	HFBC042
1320	EST02350	M85826	HFBC043
1321	EST02351	M85827	HFBC044
1322	EST02352	M85828	HFBC045
1323	EST02353	M85829	HFBC046
1324	EST02354	M85830	HFBC047
1325	EST02355	M85831	HFBC048
1326	EST02356	M85832	HFBC049
1327	EST02357	M85833	HFBC050
1328	EST02358	M85834	HFBC051
1329	EST02359	M85835	HFBC053
1330	EST02360	M85836	HFBC054
1331	EST02361	M85837	HFBC058
1332	EST02362	M85838	HFBC059
1333	EST02363	M85839	HFBC060
1334	EST02364	M85840	HFBC061
1335	EST02365	M85841	HFBC063
1336	EST02366	M85842	HFBC064
1337	EST02367	M85843	HFBC065
1338	EST02368	M85844	HFBC066
1339	EST02369	M85845	HFBC067
1340	EST02370	M85846	HFBC069
1341	EST02371	M85847	HFBC070
1342	EST02372	M85848	HFBC071
1343	EST02373	M85849	HFBC073
1344	EST02374	M85850	HFBC074
1345	EST02375	M85851	HFBC075
1346	EST02376	M85852	HFBC076
1347	EST02377	M85853	HFBC077
1348	EST02378	M85854	HFBC078
1349	EST02379	M85855	HFBC079
1350	EST02380	M85856	HFBC080
1351	EST02381	M85857	HFBC081
1352	EST02382	M85858	HFBC082
1353	EST02383	M85859	HFBC083
1354	EST02384	M85860	HFBC084
1355	EST02385	M85861	HFBC085
1356	EST02386	M85862	HFBC086
1357	EST02387	M85863	HFBC087
1358	EST02388	M85864	HFBC088
1359	EST02389	M85865	HFBC089
1360	EST02390	M85866	HFBC091
1361	EST02391	M85867	HFBC093
1362	EST02392	M85868	HFBC094
1363	EST02393	M85869	HFBC095
1364	EST02394	M85870	HFBC096
1365	EST02395	M85871	HFBCP02
1366	EST02396	M85872	HFBCP03
1367	EST02397	M85873	HFBCP04
1368	EST02398	M85874	HFBCP06
1369	EST02399	M85875	HFBCP07
1370	EST02400	M85876	HFBCP08
1371	EST02401	M85877	HFBCP09
1372	EST02402	M85878	HFBCP10
1373	EST02403	M85879	HFBCP11
1374	EST02404	M85880	HFBCP12
1375	EST02405	M85881	HFBCP13
1376	EST02406	M85882	HFBCP14
1377	EST02407	M85883	HFBCP15
1378	EST02408	M85884	HFBCP16
1379	EST02409	M85885	HFBCP17
	EST02410	M85886	HFBCP18

SEQ ID	EST#	GB#	Clone
1380	EST02411	M85887	HFBCP19
1381	EST02413	M85889	HFBCP22
1382	EST02414	M85890	HFBCP23
1383	EST02415	M85891	HFBCP49
1384	EST02416	M85892	HFBCP50
1385	EST02417	M85893	HFBCP51
1386	EST02418	M85894	HFBCP52
1387	EST02419	M85895	HFBCP54
1388	EST02421	M85897	HFBCP56
1389	EST02422	M85898	HFBCP57
1390	EST02423	M85899	HFBCP59
1391	EST02424	M85900	HFBCP60
1392	EST02425	M85901	HFBCP61
1393	EST02426	M85902	HFBCP62
1394	EST02427	M85903	HFBCP63
1395	EST02429	M85905	HFBCP66
1396	EST02430	M85906	HFBCP67
1397	EST02431	M85907	HFBCP69
1398	EST02432	M85908	HFBCP71
1399	EST02433	M85909	HFBCP72
1400	EST02434	M85910	HFBCP73
1401	EST02435	M85911	HFBCP74
1402	EST02436	M85912	HFBCP75
1403	EST02437	M85913	HFBCP76
1404	EST02438	M85914	HFBCP77
1405	EST02439	M85915	HFBCP78
1406	EST02440	M85916	HFBCP79
1407	EST02441	M85917	HFBCP81
1408	EST02443	M85918	HFBCP82
1409	EST02443	M85919	HFBCP83
1410	EST02444	M85920	HFBCP84
1411	EST02445	M85921	HFBCP85
1412	EST02446	M85922	HFBCP86
1413	EST02447	M85923	HFBCP87
1414	EST02448	M85924	HFBCP88
1415	EST02449	M85925	HFBCP89
1416	EST02450	M85926	HFBCP90
1417	EST02452	M85928	HFBCP92
1418	EST02453	M85929	HFBCP93
1419	EST02454	M85930	HFBCP94
1420	EST02456	M85932	HFBCQ14
1421	EST02457	M85933	HFBCR01
1422	EST02458	M85934	HFBCR02
1423	EST02459	M85935	HFBCR04
1424	EST02460	M85936	HFBCR05
1425	EST02461	M85937	HFBCR06
1426	EST02462	M85938	HFBCR07
1427	EST02463	M85939	HFBCR08
1428	EST02464	M85940	HFBCR09
1429	EST02465	M85941	HFBCR10
1430	EST02466	M85942	HFBCR11
1431	EST02467	M85943	HFBCR36
1432	EST02468	M85944	HFBCR37
1433	EST02469	M85945	HFBCR38
1434	EST02470	M85946	HFBCR39
1435	EST02471	M85947	HFBCR40
1436	EST02472	M85948	HFBCR42
1437	EST02473	M85949	HFBCR46
1438	EST02474	M85950	HFBCR47
1439	EST02475	M85951	HFBCR48
1440	EST02476	M85952	HFBCR49
1441	EST02477	M85953	HFBCR51
1442	EST02478	M85954	HFBCR53
1443	EST02479	M85955	HFBCR54
1444	EST02480	M85956	HFBCR55
1445	EST02481	M85957	HFBCR56

SEQ ID	EST#	GB#	Clone
1446	EST02482	M85958	HFBCR57
1447	EST02483	M85959	HFBCT01
1448	EST02484	M85960	HFBCT02
1449	EST02485	M85961	HFBCT03
1450	EST02486	M85962	HFBCT04
1451	EST02487	M85963	HFBCT05
1452	EST02488	M85964	HFBCT06
1453	EST02489	M85965	HFBCT08
1454	EST02490	M85966	HFBCT09
1455	EST02491	M85967	HFBCT10
1456	EST02492	M85968	HFBCT12
1457	EST02493	M85969	HFBCU02
1458	EST02495	M85970	HFBCU04
1459	EST02496	M85971	HFBCU06
1460	EST02497	M85972	HFBCU07
1461	EST02498	M85973	HFBCU08
1462	EST02499	M85974	HFBCU09
1463	EST02500	M85975	HFBCU11
1464	EST02501	M85976	HFBCU12
1465	EST02502	M85977	HFBCV02
1466	EST02503	M85978	HFBCV03
1467	EST02504	M85979	HFBCV04
1468	EST02505	M85980	HFBCV05
1469	EST02506	M85981	HFBCV06
1470	EST02507	M85982	HFBCV07
1471	EST02508	M85983	HFBCV08
1472	EST02509	M85984	HFBCV09
1473	EST02510	M85985	HFBCV10
1474	EST02511	M85986	HFBCV11
1475	EST02512	M85987	HFBCX02
1476	EST02513	M85988	HFBCX05
1477	EST02514	M85989	HFBCX06
1478	EST02515	M85990	HFBCX08
1479	EST02516	M85991	HFBCX10
1480	EST02517	M85992	HFBCY02
1481	EST02518	M85993	HFBCY03
1482	EST02519	M85994	HFBCY04
1483	EST02520	M85995	HFBCY05
1484	EST02521	M85996	HFBCY06
1485	EST02522	M85997	HFBCY07
1486	EST02523	M85998	HFBCY08
1487	EST02524	M85999	HFBCY09
1488	EST02525	M86000	HFBCY10
1489	EST02526	M86001	HFBCY11
1490	EST02527	M86002	HFBCY12
1491	EST02529	M86004	HFBCY15
1492	EST02530	M86005	HFBCY16
1493	EST02531	M86006	HFBCY17
1494	EST02532	M86007	HFBCY18
1495	EST02533	M86008	HFBCY19
1496	EST02534	M86009	HFBCY20
1497	EST02535	M86010	HFBCY21
1498	EST02536	M86011	HFBCY22
1499	EST02537	M86012	HFBCY24
1500	EST02538	M86013	HFBCY25
1501	EST02539	M86014	HFBCY26
1502	EST02540	M86015	HFBCY27

SEQ ID	EST#	GB#	Clone
1635	EST02682	M85316	HHCF46
1636	EST02684	M85317	HHCF93
1637	EST00798	M78648	HHCNA16
1638	EST00708	M78650	HHCNA26
1639	EST00799	M78651	HHCNA28
1640	EST00800	M78652	HHCNA29
1641	EST00801	M78653	HHCNA32
1642	EST00802	M78654	HHCNA34
1643	EST00803	M78655	HHCNA36
1644	EST01571	M77986	HHCNA37
1645	EST00804	M78656	HHCNA40
1646	EST00805	M78657	HHCNA41
1647	EST00806	M78658	HHCNA42
1648	EST00807	M78659	HHCNA45
1649	EST00808	M78660	HHCNA48
1650	EST00809	M78661	HHCNA49
1651	EST00810	M78662	HHCNA52
1652	EST00811	M78663	HHCNA53
1653	EST00812	M78664	HHCNA55
1654	EST01572	M77987	HHCNA56
1655	EST00813	M78665	HHCNA57
1656	EST00814	M78666	HHCNA58
1657	EST00815	M78667	HHCNA60
1658	EST00816	M78668	HHCNA62
1659	EST00817	M78669	HHCNA65
1660	EST00818	M78670	HHCNA68
1661	EST00819	M78671	HHCNA72
1662	EST00820	M78672	HHCNA73
1663	EST00821	M78673	HHCNA78
1664	EST00822	M78674	HHCNA80
1665	EST00823	M78675	HHCNA81
1666	EST00824	M78676	HHCNA82
1667	EST00825	M78677	HHCNA83
1668	EST00826	M78678	HHCNA84
1669	EST00827	M78679	HHCNB03
1670	EST00828	M78680	HHCNB12
1671	EST00829	M78681	HHCNB52
1672	EST00830	M78682	HHCNC04
1673	EST00831	M78683	HHCNC05
1674	EST00832	M78684	HHCNC06
1675	EST00833	M78685	HHCNC07
1676	EST00834	M78686	HHCNC09
1677	EST00835	M78687	HHCNC10
1678	EST00836	M78688	HHCNC11
1679	EST00837	M78689	HHCNC12
1680	EST00838	M78690	HHCNC13
1681	EST01573	M77988	HHCNC14
1682	EST00839	M78691	HHCNC15
1683	EST00840	M78692	HHCNC16
1684	EST00841	M78693	HHCNC17
1685	EST00842	M78694	HHCNC18
1686	EST01574	M77989	HHCNC21
1687	EST00843	M78695	HHCNC28
1688	EST00844	M78696	HHCNC36
1689	EST00845	M78697	HHCNC37
1690	EST00846	M78698	HHCNC38
1691	EST01577	M77992	HHCNC39

SEQ ID	EST#	GB#	Clone
1569	EST02612	M86087	HFBDJ16
1570	EST02613	M86088	HFBDJ17
1571	EST02614	M86089	HFBDJ19
1572	EST02615	M86090	HFBDJ20
1573	EST02616	M86091	HFBDJ21
1574	EST02617	M86092	HFBDJ22
1575	EST02618	M86093	HFBDJ24
1576	EST02619	M86094	HFBDJ25
1577	EST02620	M86095	HFBDJ29
1578	EST02621	M86096	HFBDJ31
1579	EST02622	M86097	HFBDJ32
1580	EST02623	M86098	HFBDJ35
1581	EST02624	M86099	HFBDJ38
1582	EST02625	M86100	HFBDJ40
1583	EST02626	M86102	HFBDJ42
1584	EST02629	M86103	HFBDJ43
1585	EST02630	M86104	HFBDJ44
1586	EST02631	M86105	HFBDJ45
1587	EST02632	M86106	HFBDJ46
1588	EST02633	M86107	HFBDJ47
1589	EST02634	M86108	HFBDJ49
1590	EST02635	M86109	HFBDJ50
1591	EST02636	M86110	HFBDJ51
1592	EST02637	M86111	HFBDJ52
1593	EST02638	M86112	HFBDJ53
1594	EST02639	M86113	HFBDJ54
1595	EST02640	M86114	HFBDJ55
1596	EST02641	M86115	HFBDJ57
1597	EST02642	M86116	HFBDJ61
1598	EST02643	M86117	HFBDJ62
1599	EST02644	M86118	HFBDJ63
1600	EST02645	M86119	HFBDJ64
1601	EST02646	M86120	HFBDJ65
1602	EST02647	M86121	HFBDJ67
1603	EST02648	M86122	HFBDJ69
1604	EST02649	M86123	HFBDJ70
1605	EST02650	M86124	HFBDJ71
1606	EST02651	M86125	HFBDJ72
1607	EST02652	M86126	HFBDJ74
1608	EST02653	M86127	HFBDJ75
1609	EST02654	M86128	HFBDJ76
1610	EST02655	M86129	HFBDJ77
1611	EST02656	M86130	HFBDJ79
1612	EST02657	M86131	HFBDJ80
1613	EST02658	M86132	HFBDJ86
1614	EST02659	M86133	HFBDJ87
1615	EST02660	M86134	HFBDJ89
1616	EST02661	M86135	HFBDJ91
1617	EST02662	M86136	HFBDJ92
1618	EST02663	M86137	HFBDJ93
1619	EST02664	M86138	HFBDK02
1620	EST02665	M86139	HFBDK11
1621	EST02667	M86141	HHCE25
1622	EST02668	M86142	HHCE27
1623	EST02669	M86143	HHCE30
1624	EST02670	M86144	HHCE32
1625	EST02672	M86145	HHCE36
1626	EST02673	M86146	HHCE39
1627	EST02674	M86147	HHCF07
1628	EST02675	M86148	HHCF07
1629	EST02677	M85314	HHCF19
1630	EST02678	M86149	HHCF19
1631	EST02679	M86150	HHCF40
1632	EST02680	M86151	HHCF44
1633	EST02681	M85315	HHCF44
1634		M86152	HHCF46

SEQ ID	EST#	GB#	Clone
1503	EST02541	M86016	HFBCY28
1504	EST02542	M86017	HFBCY30
1505	EST02543	M86018	HFBCY31
1506	EST02545	M86020	HFBCY33
1507	EST02546	M86021	HFBCY34
1508	EST02547	M86022	HFBCY35
1509	EST02548	M86023	HFBCY36
1510	EST02549	M86024	HFBCY37
1511	EST02550	M86025	HFBCY38
1512	EST02551	M86026	HFBCY40
1513	EST02552	M86027	HFBCY41
1514	EST02553	M86028	HFBCY43
1515	EST02554	M86029	HFBCY44
1516	EST02555	M86030	HFBCY45
1517	EST02558	M86033	HFBCY50
1518	EST02559	M86034	HFBCY51
1519	EST02560	M86035	HFBCY52
1520	EST02561	M86036	HFBCY53
1521	EST02562	M86037	HFBCY54
1522	EST02563	M86038	HFBCY55
1523	EST02564	M86039	HFBCZ01
1524	EST02565	M86040	HFBCZ04
1525	EST02566	M86041	HFBCZ05
1526	EST02567	M86042	HFBCZ06
1527	EST02568	M86043	HFBCZ07
1528	EST02569	M86044	HFBCZ08
1529	EST02570	M86045	HFBCZ09
1530	EST02571	M86046	HFBCZ10
1531	EST02572	M86047	HFBCZ11
1532	EST02573	M86048	HFBCZ12
1533	EST02574	M86049	HFBDE01
1534	EST02575	M86050	HFBDE03
1535	EST02576	M86051	HFBDE04
1536	EST02577	M86052	HFBDE05
1537	EST02578	M86053	HFBDE06
1538	EST02579	M86054	HFBDE07
1539	EST02580	M86055	HFBDE09
1540	EST02581	M86056	HFBDE12
1541	EST02582	M86057	HFBDF01
1542	EST02583	M86058	HFBDF02
1543	EST02584	M86059	HFBDF03
1544	EST02585	M86061	HFBDF04
1545	EST02586	M86062	HFBDF06
1546	EST02587	M86063	HFBDF10
1547	EST02588	M86064	HFBDF11
1548	EST02589	M86065	HFBDI01
1549	EST02590	M86066	HFBDI02
1550	EST02591	M86067	HFBDI03
1551	EST02592	M86068	HFBDI04
1552	EST02593	M86069	HFBDI05
1553	EST02594	M86070	HFBDI06
1554	EST02595	M86072	HFBDI07
1555	EST02597	M86073	HFBDI10
1556	EST02598	M86074	HFBDI11
1557	EST02599	M86075	HFBDI12
1558	EST02600	M86076	HFBDJ01
1559	EST02601	M86077	HFBDJ02
1560	EST02602	M86078	HFBDJ04
1561	EST02603	M86079	HFBDJ06
1562	EST02604	M86080	HFBDJ07
1563	EST02605	M86081	HFBDJ09
1564	EST02606	M86082	HFBDJ10
1565	EST02607	M86083	HFBDJ11
1566	EST02608	M86084	HFBDJ12
1567	EST02609	M86085	HFBDJ13
1568	EST02610	M86086	HFBDJ14
	EST02611		HFBDJ15

SEQ ID	EST#	GB#	Clone
1824	EST01617	M78030	HHCMG04
1825	EST00955	M78807	HHCMG07
1826	EST00956	M78808	HHCMG08
1827	EST01618	M78031	HHCMG10
1828	EST00957	M78809	HHCMG11
1829	EST00958	M78810	HHCMG12
1830	EST00959	M78811	HHCMG13
1831	EST01619	M78032	HHCMG15
1832	EST00960	M78812	HHCMG16
1833	EST00961	M78813	HHCMG17
1834	EST01620	M78033	HHCMG18
1835	EST00962	M78814	HHCMG21
1836	EST01622	M78035	HHCMG22
1837	EST00963	M78815	HHCMG25
1838	EST00964	M78816	HHCMG27
1839	EST00965	M78817	HHCMG28
1840	EST00966	M78818	HHCMG30
1841	EST00967	M78819	HHCMG31
1842	EST00968	M78820	HHCMG36
1843	EST00969	M78821	HHCMG38
1844	EST00970	M78822	HHCMG40
1845	EST00971	M78823	HHCMG44
1846	EST00972	M78824	HHCMG45
1847	EST00973	M78825	HHCMG51
1848	EST00974	M78826	HHCMG53
1849	EST00975	M78827	HHCMG55
1850	EST01624	M78037	HHCMG63
1851	EST00976	M78828	HHCMG71
1852	EST01625	M78038	HHCMG75
1853	EST00977	M78829	HHCMG79
1854	EST00978	M78830	HHCMG81
1855	EST00979	M78831	HHCMG83
1856	EST01627	M78040	HHCMG86
1857	EST00980	M78832	HHCMG87
1858	EST00981	M78833	HHCMG88
1859	EST00982	M78834	HHCMG89
1860	EST00983	M78835	HHCMG91
1861	EST00984	M78836	HHCMG92
1862	EST00985	M78837	HHCMG94
1863	EST00986	M78838	HHCMG95
1864	EST00987	M78839	HHCMG96
1865	EST00988	M78840	HHCMH02
1866	EST00989	M78841	HHCMH08
1867	EST00990	M78842	HHCMH14
1868	EST00991	M78843	HHCMH17
1869	EST00992	M78844	HHCMH17
1870	EST00993	M78845	HHCMH18
1871	EST00994	M78846	HHCMH22
1872	EST00995	M78847	HHCMH23
1873	EST01630	M78043	HHCMH25
1874	EST00996	M78848	HHCMH27
1875	EST01631	M78044	HHCMH29
1876	EST00997	M78849	HHCMH31
1877	EST00998	M78850	HHCMH33
1878	EST00999	M78851	HHCMH40
1879	EST01633	M78046	HHCMH42
1880	EST01634	M78047	

SEQ ID	EST#	GB#	Clone
1758	EST00901	M78753	HHCME17
1759	EST00902	M78754	HHCME22
1760	EST00903	M78755	HHCME23
1761	EST01598	M78012	HHCME24
1762	EST00904	M78756	HHCME25
1763	EST00905	M78757	HHCME26
1764	EST02690	M86158	HHCME27
1765	EST01600	M78014	HHCME29
1766	EST00906	M78758	HHCME30
1767	EST00907	M78759	HHCME31
1768	EST00908	M78760	HHCME33
1769	EST00909	M78761	HHCME35
1770	EST00910	M78762	HHCME37
1771	EST01601	M78015	HHCME40
1772	EST02691	M86159	HHCME45
1773	EST00911	M78763	HHCME47
1774	EST00912	M78764	HHCME48
1775	EST02692	M86160	HHCME52
1776	EST01603	M78017	HHCME53
1777	EST00913	M78765	HHCME54
1778	EST00914	M78766	HHCME56
1779	EST00915	M78767	HHCME57
1780	EST00916	M78768	HHCME58
1781	EST00917	M78769	HHCME60
1782	EST00918	M78770	HHCME61
1783	EST00919	M78771	HHCME65
1784	EST00920	M78772	HHCME68
1785	EST00921	M78773	HHCME69
1786	EST00922	M78774	HHCME71
1787	EST00923	M78775	HHCME73
1788	EST00924	M78776	HHCME78
1789	EST00925	M78777	HHCME90
1790	EST00926	M78778	HHCME91
1791	EST00927	M78779	HHCMF04
1792	EST00928	M78780	HHCMF05
1793	EST00929	M78781	HHCMF09
1794	EST01607	M78021	HHCMF13
1795	EST00930	M78782	HHCMF14
1796	EST00931	M78783	HHCMF29
1797	EST00932	M78784	HHCMF34
1798	EST00933	M78785	HHCMF36
1799	EST00934	M78786	HHCMF37
1800	EST00935	M78787	HHCMF41
1801	EST00936	M78788	HHCMF44
1802	EST00937	M78789	HHCMF47
1803	EST01613	M78026	HHCMF48
1804	EST00938	M78790	HHCMF50
1805	EST00939	M78791	HHCMF52
1806	EST00940	M78792	HHCMF53
1807	EST00941	M78793	HHCMF57
1808	EST00942	M78794	HHCMF63
1809	EST00943	M78795	HHCMF65
1810	EST00944	M78796	HHCMF66
1811	EST00945	M78797	HHCMF69
1812	EST02693	M86161	HHCMF70
1813	EST00946	M78798	HHCMF73
1814	EST00947	M78799	HHCMF76
1815	EST01615	M78028	HHCMF77
1816	EST00948	M78800	HHCMH80
1817	EST00949	M78801	HHCMH83
1818	EST01616	M78029	HHCMH86
1819	EST00950	M78802	HHCMH93
1820	EST00951	M78803	HHCMH96
1821	EST00952	M78804	HHCMG02
1822	EST00953	M78805	HHCMG03
1823	EST00954	M78806	

SEQ ID	EST#	GB#	Clone
1692	EST00847	M78699	HHCMC42
1693	EST00848	M78700	HHCMC44
1694	EST00849	M78701	HHCMC47
1695	EST00850	M78702	HHCMC52
1696	EST00851	M78703	HHCMC53
1697	EST00852	M78704	HHCMC54
1698	EST02685	M86154	HHCMC55
1699	EST01579	M86155	HHCMC60
1700	EST00853	M77994	HHCMC63
1701	EST00854	M78705	HHCMC70
1702	EST00855	M78706	HHCMC71
1703	EST01580	M78707	HHCMC72
1704	EST00856	M77995	HHCMC77
1705	EST00857	M78708	HHCMC78
1706	EST01581	M78709	HHCMC79
1707	EST00858	M77996	HHCMC82
1708	EST00859	M78710	HHCMC85
1709	EST00860	M78711	HHCMC87
1710	EST00861	M78712	HHCMC87
1711	EST00862	M78713	HHCMC88
1712	EST00863	M78714	HHCMD02
1713	EST00864	M78715	HHCMD04
1714	EST01583	M78716	HHCMD06
1715	EST00865	M77998	HHCMD08
1716	EST00866	M78717	HHCMD10
1717	EST00867	M78718	HHCMD15
1718	EST00868	M78719	HHCMD16
1719	EST00869	M78720	HHCMD17
1720	EST00870	M78721	HHCMD20
1721	EST00871	M78722	HHCMD21
1722	EST00872	M78723	HHCMD22
1723	EST00873	M78724	HHCMD27
1724	EST00874	M78725	HHCMD28
1725	EST01588	M78726	HHCMD29
1726	EST00875	M78003	HHCMD31
1727	EST00876	M78727	HHCMD35
1728	EST00877	M78728	HHCMD36
1729	EST00878	M78729	HHCMD37
1730	EST00879	M78730	HHCMD38
1731	EST01590	M78731	HHCMD39
1732	EST01591	M78005	HHCMD40
1733	EST00880	M78006	HHCMD44
1734	EST00881	M78732	HHCMD49
1735	EST01592	M78733	HHCMD50
1736	EST00882	M78007	HHCMD53
1737	EST00883	M78734	HHCMD54
1738	EST00885	M78735	HHCMD76
1739	EST02687	M86156	HHCMD77
1740	EST00886	M78736	HHCMD79
1741	EST00887	M78739	HHCMD80
1742	EST00888	M78740	HHCMD81
1743	EST00889	M78741	HHCMD83
1744	EST00890	M78742	HHCMD84
1745	EST00891	M78743	HHCMD85
1746	EST00892	M78744	HHCMD87
1747	EST00893	M78745	HHCMD88
1748	EST01593	M78008	HHCMD90
1749	EST00894	M78746	HHCMD93
1750	EST00895	M78747	HHCMD94
1751	EST00896	M78748	HHCME06
1752	EST00897	M78749	HHCME07
1753	EST00898	M78750	HHCME10
1754	EST00899	M78751	HHCME11
1755	EST00900	M78752	HHCME13
1756	EST01594	M78009	
1757			

SEQ ID	EST#	GB#	Clone
2013	EST01107	M78959	HHCPE80
2014	EST01108	M78960	HHCPE81
2015	EST01109	M78961	HHCPE83
2016	EST01110	M78962	HHCPE85
2017	EST01111	M78963	HHCPE89
2018	EST01112	M78964	HHCPE94
2019	EST01113	M78965	HHCPE96
2020	EST01114	M78966	HHCPF04
2021	EST01115	M78967	HHCPF10
2022	EST01116	M78968	HHCPF11
2023	EST01117	M78969	HHCPF12
2024	EST01118	M78970	HHCPF16
2025	EST01119	M78971	HHCPF24
2026	EST01120	M78972	HHCPF29
2027	EST01121	M78973	HHCPF33
2028	EST01122	M78974	HHCPF43
2029	EST01123	M78975	HHCPF44
2030	EST01124	M78976	HHCPF50
2031	EST01125	M78977	HHCPF52
2032	EST01126	M78978	HHCPF54
2033	EST01127	M78979	HHCPF56
2034	EST01128	M78980	HHCPF58
2035	EST01129	M78981	HHCPF61
2036	EST01130	M78982	HHCPF62
2037	EST01131	M78983	HHCPF66
2038	EST01132	M78984	HHCPF72
2039	EST01133	M78985	HHCPF79
2040	EST01134	M78986	HHCPF81
2041	EST01135	M78987	HHCPF84
2042	EST01136	M78988	HHCPF86
2043	EST01137	M78989	HHCPF87
2044	EST01138	M78990	HHCPF91
2045	EST01139	M78991	HHCPF92
2046	EST01140	M78992	HHCPG03
2047	EST01141	M78993	HHCPG05
2048	EST01142	M78994	HHCPG13
2049	EST01143	M78995	HHCPG14
2050	EST01144	M78996	HHCPG35
2051	EST01145	M78997	HHCPG37
2052	EST01146	M78998	HHCPG39
2053	EST01147	M78999	HHCPG42
2054	EST01148	M79000	HHCPG45
2055	EST01149	M79001	HHCPG47
2056	EST01150	M79002	HHCPG48
2057	EST01151	M79003	HHCPG62
2058	EST01152	M79004	HHCPG68
2059	EST01153	M79005	HHCPG70
2060	EST01154	M79006	HHCPG72
2061	EST01155	M79007	HHCPG76
2062	EST01156	M79008	HHCPG77
2063	EST01157	M79009	HHCPG82
2064	EST01158	M79010	HHCPG83
2065	EST01159	M79011	HHCPG88
2066	EST01160	M79012	HHCPG96
2067	EST01161	M79013	HHCPH13
2068	EST01162	M79014	HHCPH18
2069	EST01163	M79015	HHCPH19

SEQ ID	EST#	GB#	Clone
1947	EST01052	M78904	HHCPD05
1948	EST01053	M78905	HHCPD05
1949	EST01054	M78906	HHCPD08
1950	EST01055	M78907	HHCPD08
1951	EST01056	M78908	HHCPD10
1952	EST01057	M78909	HHCPD11
1953	EST01058	M78910	HHCPD14
1954	EST01059	M78911	HHCPD16
1955	EST01060	M78912	HHCPD18
1956	EST01061	M78913	HHCPD29
1957	EST01062	M86166	HHCPD37
1958	EST01040	M78914	HHCPD40
1959	EST01041	M78915	HHCPD42
1960	EST02697	M78916	HHCPD44
1961	EST02698	M78917	HHCPD45
1962	EST01042	M78918	HHCPD49
1963	EST01043	M78919	HHCPD52
1964	EST01044	M78920	HHCPD53
1965	EST01045	M78921	HHCPD55
1966	EST01046	M78922	HHCPD58
1967	EST01047	M78923	HHCPD60
1968	EST01048	M78924	HHCPD62
1969	EST01049	M78925	HHCPD65
1970	EST01050	M78926	HHCPD66
1971	EST01051	M78927	HHCPD20
1972	EST01070	M78928	HHCPD26
1973	EST01071	M78929	HHCPD27
1974	EST01072	M78930	HHCPD28
1975	EST01073	M78931	HHCPD34
1976	EST01074	M78932	HHCPD43
1977	EST01075	M78933	HHCPD48
1978	EST01076	M78934	HHCPD49
1979	EST01077	M78935	HHCPD50
1980	EST01078	M78936	HHCPD56
1981	EST01079	M86168	HHCPD59
1982	EST01080	M78937	HHCPD70
1983	EST01081	M78938	HHCPD76
1984	EST01082	M78939	HHCPD80
1985	EST01083	M78940	HHCPE09
1986	EST01084	M78941	HHCPE12
1987	EST02700	M78942	HHCPE14
1988	EST01085	M78943	HHCPE16
1989	EST01086	M78944	HHCPE17
1990	EST01087	M78945	HHCPE26
1991	EST01088	M78946	HHCPE27
1992	EST01089	M78947	HHCPE29
1993	EST01090	M78948	HHCPE30
1994	EST01091	M78949	HHCPE33
1995	EST01092	M78950	HHCPE37
1996	EST01093	M78951	HHCPE38
1997	EST01094	M78085	HHCPE69
1998	EST01095	M78952	HHCPE70
1999	EST01096	M78953	HHCPE71
2000	EST01097	M78954	HHCPE72
2001	EST01098	M78955	HHCPE73
2002	EST01099	M78087	HHCPE74
2003	EST01675	M78956	HHCPE75
2004	EST01676	M78957	HHCPE76
2005	EST01100	M78958	HHCPE78
2006	EST01101		HHCPE79
2007	EST01102		
2008	EST01103		
2009	EST01677		
2010	EST01104		
2011	EST01105		
2012	EST01106		

SEQ ID	EST#	GB#	Clone
1881	EST01000	M78852	HHCNH48
1882	EST01638	M78851	HHCNH53
1883	EST01001	M78853	HHCNH54
1884	EST01002	M78854	HHCNH55
1885	EST01639	M78855	HHCNH61
1886	EST01003	M78856	HHCNH69
1887	EST01004	M78857	HHCNH72
1888	EST01005	M78858	HHCNH73
1889	EST01006	M78859	HHCNH74
1890	EST01007	M78860	HHCNH75
1891	EST01008	M78861	HHCNH77
1892	EST01009	M78055	HHCNH81
1893	EST01642	M78056	HHCNH89
1894	EST01643	M78862	HHCNH89
1895	EST01010	M78863	HHCNH91
1896	EST01011	M78864	HHCPA01
1897	EST01012	M78865	HHCPA05
1898	EST01013	M78866	HHCPA14
1899	EST01014	M78867	HHCPA16
1900	EST01015	M78868	HHCPA17
1901	EST01016	M78869	HHCPA18
1902	EST01017	M78870	HHCPA25
1903	EST01018	M78871	HHCPA26
1904	EST01019	M78872	HHCPA28
1905	EST01020	M78873	HHCPA29
1906	EST01021	M78874	HHCPA42
1907	EST01022	M78875	HHCPA49
1908	EST01023	M78876	HHCPA52
1909	EST01024	M78058	HHCPA59
1910	EST01645	M86162	HHCPA59
1911	EST02694	M78877	HHCPA60
1912	EST01025	M78878	HHCPA67
1913	EST01646	M78059	HHCPA76
1914	EST01026	M78879	HHCPA78
1915	EST01027	M78880	HHCPA81
1916	EST01028	M78881	HHCPA83
1917	EST01029	M86163	HHCPA83
1918	EST02695	M78882	HHCPA88
1919	EST01030	M78883	HHCPA89
1920	EST01031	M78060	HHCPA90
1921	EST01647	M78884	HHCPB01
1922	EST01032	M78885	HHCPB10
1923	EST01033	M78886	HHCPB11
1924	EST01034	M78887	HHCPB15
1925	EST01035	M78888	HHCPB19
1926	EST01036	M78889	HHCPB21
1927	EST01037	M78890	HHCPB22
1928	EST01038	M78891	HHCPB27
1929	EST01039	M78892	HHCPB28
1930	EST01040	M78893	HHCPB34
1931	EST01041	M78894	HHCPB34
1932	EST01042	M78063	HHCPB41
1933	EST01650	M78895	HHCPB49
1934	EST01043	M78896	HHCPB52
1935	EST01044	M78897	HHCPB57
1936	EST01045	M78065	HHCPB58
1937	EST01652	M78067	HHCPB60
1938	EST01654	M78068	HHCPB63
1939	EST01655	M78898	HHCPB66
1940	EST01046	M78899	HHCPB69
1941	EST01047	M78900	HHCPB82
1942	EST01048	M78901	HHCPB85
1943	EST01049	M78902	HHCPB90
1944	EST01050	M78903	HHCPB96
1945	EST01051	M86164	HHCPC02
1946	EST02696		HHCPC03

SEQ ID	EST#	GB#	Clone
2202	EST01732	M78141	HHCPL63
2203	EST01264	M79116	HHCPL65
2204	EST01265	M79117	HHCPL76
2205	EST01735	M78144	HHCPL77
2206	EST01736	M78145	HHCPL79
2207	EST01266	M79118	HHCPL81
2208	EST01267	M79119	HHCPL82
2209	EST02717		HHCPL84
2210	EST01268	M79120	HHCPL87
2211	EST01269	M79121	HHCPL93
2212	EST01270	M79122	HHCPL96
2213	EST01271	M79123	HHCPM02
2214	EST01272	M79124	HHCPM03
2215	EST01273	M79125	HHCPM06
2216	EST01737	M78146	HHCPM10
2217	EST01738	M78147	HHCPM17
2218	EST01274	M79126	HHCPM22
2219	EST01275	M79127	HHCPM24
2220	EST01740	M78148	HHCPM28
2221	EST01741	M78149	HHCPM30
2222	EST01276	M79128	HHCPM31
2223	EST01742	M78150	HHCPM37
2224	EST01277	M79129	HHCPM42
2225	EST01278	M79130	HHCPM44
2226	EST01744	M78152	HHCPM47
2227	EST01279	M79131	HHCPM49
2228	EST01280	M79132	HHCPM50
2229	EST01281	M79133	HHCPM58
2230	EST01282	M79134	HHCPM59
2231	EST01746	M78154	HHCPM60
2232	EST01283	M79135	HHCPM61
2233	EST01284	M79136	HHCPM65
2234	EST01285	M79137	HHCPM66
2235	EST01286	M79138	HHCPM70
2236	EST01287	M79139	HHCPM72
2237	EST01288	M79140	HHCPM75
2238	EST01289	M79141	HHCPM77
2239	EST01290	M79142	HHCPM81
2240	EST01291	M79143	HHCPM82
2241	EST01747	M78155	HHCPM83
2242	EST01292	M79144	HHCPN04
2243	EST01293	M79145	HHCPN06
2244	EST01294	M79146	HHCPN09
2245	EST01748	M78156	HHCPN14
2246	EST01295	M79147	HHCPN15
2247	EST01296	M79148	HHCPN18
2248	EST01297	M79149	HHCPN24
2249	EST01298	M79150	HHCPN25
2250	EST01299	M79151	HHCPN26
2251	EST01300	M79152	HHCPN30
2252	EST01750	M78158	HHCPN31
2253	EST01301	M79153	HHCPN32
2254	EST01751	M78159	HHCPN36
2255	EST01302	M79154	HHCPN38
2256	EST02718		HHCPN46
2257	EST01303	M79155	HHCPN48
2258	EST01754	M78161	HHCPN50

SEQ ID	EST#	GB#	Clone
2136	EST01210	M79062	HHCPJ33
2137	EST01211	M79063	HHCPJ40
2138	EST01715	M78124	HHCPJ42
2139	EST01212	M79064	HHCPJ44
2140	EST01213	M79065	HHCPJ50
2141	EST01214	M79066	HHCPJ54
2142	EST01215	M79067	HHCPJ60
2143	EST01216	M79068	HHCPJ62
2144	EST01217	M79069	HHCPJ65
2145	EST01218	M79070	HHCPJ69
2146	EST01219	M79071	HHCPJ76
2147	EST01220	M79072	HHCPJ77
2148	EST01221	M79073	HHCPJ80
2149	EST01222	M79074	HHCPJ81
2150	EST01223	M79075	HHCPJ82
2151	EST01224	M79076	HHCPJ83
2152	EST01225	M79077	HHCPJ91
2153	EST01226	M79078	HHCPJ96
2154	EST01227	M79079	HHCPK02
2155	EST01718	M78127	HHCPK05
2156	EST01719	M78128	HHCPK09
2157	EST01228	M79080	HHCPK10
2158	EST01229	M79081	HHCPK17
2159	EST01230	M79082	HHCPK20
2160	EST01231	M79083	HHCPK21
2161	EST01232	M79084	HHCPK26
2162	EST01233	M79085	HHCPK27
2163	EST01234	M79086	HHCPK31
2164	EST01720	M78129	HHCPK38
2165	EST01236	M79088	HHCPK40
2166	EST01237	M79089	HHCPK42
2167	EST01238	M79090	HHCPK45
2168	EST01722	M78131	HHCPK49
2169	EST01239	M79091	HHCPK55
2170	EST01240	M79092	HHCPK66
2171	EST01241	M79093	HHCPK67
2172	EST01724	M78133	HHCPK68
2173	EST01242	M79094	HHCPK77
2174	EST01243	M79095	HHCPK82
2175	EST01244	M79096	HHCPK85
2176	EST01245	M79097	HHCPK88
2177	EST01726	M78135	HHCPK91
2178	EST01246	M79098	HHCPL02
2179	EST01247	M79099	HHCPL03
2180	EST01248	M79100	HHCPL05
2181	EST01249	M79101	HHCPL06
2182	EST01250	M79102	HHCPL07
2183	EST01251	M79103	HHCPL08
2184	EST01252	M79104	HHCPL17
2185	EST01253	M79105	HHCPL19
2186	EST01727	M78136	HHCPL20
2187	EST01254	M79106	HHCPL21
2188	EST01255	M79107	HHCPL25
2189	EST01728	M78137	HHCPL27
2190	EST01256	M79108	HHCPL32
2191	EST01257	M79109	HHCPL36
2192	EST01258	M79110	HHCPL37
2193	EST01729	M78138	HHCPL39
2194	EST01259	M79111	HHCPL48
2195	EST01260	M79112	HHCPL51
2196	EST01261	M79113	HHCPL52
2197	EST01730	M78139	HHCPL53
2198	EST01262	M79114	HHCPL54
2199	EST01731	M78140	HHCPL58
2200	EST01263	M79115	HHCPL60

SEQ ID	EST#	GB#	Clone
2070	EST01151	M79003	HHCPH24
2071	EST01696	M78106	HHCPH32
2072	EST01152	M79004	HHCPH33
2073	EST01697	M78107	HHCPH35
2074	EST01698	M78108	HHCPH39
2075	EST01153	M79005	HHCPH40
2076	EST02702	M86170	HHCPH42
2077	EST01154	M79006	HHCPH43
2078	EST01155	M79007	HHCPH44
2079	EST01156	M79008	HHCPH45
2080	EST01157	M79009	HHCPH56
2081	EST01158	M79010	HHCPH59
2082	EST01159	M79011	HHCPH60
2083	EST01160	M79012	HHCPH61
2084	EST01161	M79013	HHCPH63
2085	EST01162	M79014	HHCPH68
2086	EST01163	M79015	HHCPH70
2087	EST01164	M79016	HHCPH72
2088	EST01166	M78109	HHCPH75
2089	EST01699	M79019	HHCPH79
2090	EST01167	M79020	HHCPH80
2091	EST01168	M78110	HHCPH84
2092	EST01700	M79022	HHCPH85
2093	EST01170	M79023	HHCPI01
2094	EST01171	M78111	HHCPI03
2095	EST01701	M79024	HHCPI05
2096	EST01172	M79025	HHCPI08
2097	EST01173	M79026	HHCPI09
2098	EST01174	M79027	HHCPI20
2099	EST01175	M79028	HHCPI28
2100	EST01176	M79029	HHCPI34
2101	EST01177	M79030	HHCPI38
2102	EST01178	M79031	HHCPI38
2103	EST01179	M79032	HHCPI40
2104	EST01180	M79033	HHCPI41
2105	EST01181	M79034	HHCPI41
2106	EST01182	M79035	HHCPI50
2107	EST01183	M79036	HHCPI56
2108	EST01184	M79037	HHCPI56
2109	EST01185	M79038	HHCPI58
2110	EST01186	M79039	HHCPI58
2111	EST01187	M79040	HHCPI63
2112	EST01188	M79041	HHCPI67
2113	EST01189	M79042	HHCPI68
2114	EST01190	M79043	HHCPI70
2115	EST01191	M79044	HHCPI74
2116	EST01192	M79045	HHCPI77
2117	EST01193	M79046	HHCPI84
2118	EST01194	M79047	HHCPI85
2119	EST01195	M79048	HHCPI88
2120	EST01196	M78120	HHCPI90
2121	EST01711	M79049	HHCPI94
2122	EST01197	M78122	HHCPI96
2123	EST01713	M79050	HHCPJ03
2124	EST01198	M79051	HHCPJ06
2125	EST01199	M79052	HHCPJ09
2126	EST01200	M79053	HHCPJ12
2127	EST01201	M79054	HHCPJ16
2128	EST01202	M79055	HHCPJ20
2129	EST01203	M79056	HHCPJ21
2130	EST01204	M79057	HHCPJ22
2131	EST01205	M79058	HHCPJ23
2132	EST01206	M79059	HHCPJ27
2133	EST01207	M79060	HHCPJ29
2134	EST01208	M79061	HHCPJ32
2135	EST01209		

SEQ ID	EST#	GB#	Clone
2392	EST01416	M79241	HRBAA06
2393	EST01417	M79242	HRBAA07
2394	EST01418	M79243	HRBAA20
2395	EST01419	M79244	HRBAA26
2396	EST01420	M79245	HRBAA27
2397	EST01421	M79246	HRBBA04
2398	EST01422	M79247	HRBBA06
2399	EST01423	M79248	HRBBA08
2400	EST01424	M78227	HRBBA11
2401	EST01424	M78227	HRBBA14
2402	EST01826	M78269	HRBBA21
2403	EST01425	M79270	HRBBA22
2404	EST01426	M79271	HRBBA23
2405	EST01427	M79272	HRBBA24
2406	EST02713	M86178	HRHAA22
2407	EST02714	M86179	HRHAA23
2408	EST00244	M62182	HHCG50
2409	EST00273	M62211	HFBA76

SEQ ID	EST#	GB#	Clone
2326	EST01354	M78198	HHDP006
2327	EST01355	M79206	HHDP007
2328	EST01792	M79207	HHDP012
2329	EST01793	M78199	HHDP013
2330	EST01356	M78200	HHDP016
2331	EST01794	M78201	HHDP022
2332	EST01357	M79209	HHDP023
2333	EST01358	M79210	HHDP029
2334	EST01359	M79211	HHDP050
2335	EST01360	M79212	HHDP060
2336	EST01361	M79213	HHDP062
2337	EST02706	M86174	HHDP068
2338	EST01362	M79214	HHDP070
2339	EST01802	M78209	HHDP075
2340	EST01344	M79216	HHDP088
2341	EST01345	M79217	HHDP089
2342	EST01346	M79218	HHDP090
2343	EST01347	M79219	HHDP091
2344	EST01348	M79220	HHDP092
2345	EST01349	M79221	HHDP094
2346	EST01370	M79222	HHDP095
2347	EST01371	M79223	HHDP102
2348	EST01372	M79224	HHDR10
2349	EST02708	M86176	HHDR11
2350	EST01373	M79225	HHDR13
2351	EST01374	M79226	HHDR15
2352	EST01806	M78213	HHDR16
2353	EST01375	M79227	HHDR19
2354	EST01376	M79228	HHDR22
2355	EST01377	M79229	HHDR23
2356	EST01378	M79230	HHDR28
2357	EST01379	M79231	HHDR49
2358	EST01380	M79232	HHDR51
2359	EST01381	M79233	HHDR52
2360	EST01382	M79234	HHDR53
2361	EST01383	M79235	HHDR62
2362	EST01384	M79236	HHDR66
2363	EST01385	M79237	HHDR68
2364	EST01386	M79238	HHDR75
2365	EST01387	M79239	HHDR78
2366	EST01388	M79240	HHDR86
2367	EST01389	M79241	HHDR90
2368	EST01811	M78218	HHDR95
2369	EST01390	M79243	HHDR96
2370	EST01391	M79244	HHDS02
2371	EST01392	M79245	HHDS06
2372	EST01393	M79246	HHDS12
2373	EST01394	M78222	HHDS17
2374	EST01815	M79247	HHDS18
2375	EST01395	M79248	HHDS29
2376	EST01396	M79249	HHDS30
2377	EST01397	M79250	HHDS35
2378	EST01398	M79251	HHDS36
2379	EST01399	M79252	HHDS39
2380	EST01400	M79253	HHDS40
2381	EST01401	M79254	HHDS43
2382	EST01402	M79255	HHDS44
2383	EST01403	M79256	HHDS51
2384	EST01816	M78223	HHDS56
2385	EST01404	M79257	HHDS60
2386	EST01405	M79258	HHDS62
2387	EST01406	M79259	HHDS63
2388	EST02712	M86177	HHDS71
2389	EST01407	M79260	HHDS78
2390	EST01415	M79260	HHDS95
2391			HRBAA06

SEQ ID	EST#	GB#	Clone
2259	EST01304	M79156	HHCPN52
2260	EST01305	M79157	HHCPN53
2261	EST01755	M78162	HHCPN60
2262	EST01306	M79158	HHCPN63
2263	EST01307	M79159	HHCPN64
2264	EST01308	M79160	HHCPN65
2265	EST01309	M79161	HHCPN67
2266	EST01310	M78163	HHCPN70
2267	EST01756	M78163	HHCPN76
2268	EST01311	M79164	HHCPN92
2269	EST01312	M79165	HHCPN96
2270	EST01313	M79166	HHCPO01
2271	EST01314	M78169	HHCPO03
2272	EST01762	M79167	HHCPO05
2273	EST01315	M86172	HHCPO06
2274	EST02704	M79168	HHCPO07
2275	EST01316	M79169	HHCPO08
2276	EST01317	M79170	HHCPO10
2277	EST01318	M79171	HHCPO19
2278	EST01319	M79172	HHCPO21
2279	EST01320	M78170	HHCPO22
2280	EST01763	M78173	HHCPO24
2281	EST01321	M78171	HHCPO25
2282	EST01764	M79174	HHCPO45
2283	EST01322	M79175	HHCPO51
2284	EST01323	M78175	HHCPO57
2285	EST01768	M78177	HHCPO64
2286	EST01770	M78176	HHCPO65
2287	EST01324	M79176	HHCPO72
2288	EST01325	M79177	HHCPO74
2289	EST01772	M78179	HHCPO75
2290	EST01773	M78180	HHCPO76
2291	EST01326	M79178	HHCPO81
2292	EST01327	M79179	HHCPO82
2293	EST01328	M79180	HHCPO83
2294	EST01329	M79181	HHCPO86
2295	EST01330	M78182	HHCPO87
2296	EST01775	M78183	HHCPO88
2297	EST01331	M79183	HHCPO92
2298	EST01332	M79184	HHCPP03
2299	EST01333	M79185	HHCPP04
2300	EST01334	M78186	HHCPP05
2301	EST01779	M78184	HHCPP06
2302	EST01335	M78187	HHCPP07
2303	EST01780	M78187	HHCPP12
2304	EST01336	M79188	HHCPP13
2305	EST01337	M79189	HHCPP15
2306	EST02705	M86173	HHCPP21
2307	EST01339	M79191	HHCPP22
2308	EST01340	M79192	HHCPP27
2309	EST01341	M79193	HHCPP37
2310	EST01342	M79194	HHCPP39
2311	EST01343	M79195	HHCPP41
2312	EST01344	M79196	HHCPP45
2313	EST01345	M79197	HHCPP47
2314	EST01346	M78188	HHCPP51
2315	EST01782	M79198	HHCPP53
2316	EST01347	M79199	HHCPP56
2317	EST01348	M79200	HHCPP64
2318	EST01349	M79201	HHCPP66
2319	EST01784	M78191	HHCPP69
2320	EST01350	M79202	HHCPP75
2321	EST01351	M79203	HHCPP76
2322	EST01789	M78196	HHCPP77
2323	EST01352	M79204	HHCPP81
2324	EST01353	M79205	HHCPP96
2325			HHCP005

NOTE REGARDING SEQUENCE LISTINGS: The listings of SEQ ID NOS: 1-2421 are in numerical order. However, an occasional number (for example, SEQ ID NO: 44) is not found in this list. In all, 9 SEQ ID NOS are not used. Nevertheless, the convention "1-2421" is used, for example, to refer to all the SEQ ID NOS in the following list, while "1-315" is used, for example, to refer to all the listed sequences falling between SEQ ID NO 1 and SEQ ID NO 315.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Venter, J. Craig
Adams, Mark D.
Moreno, Ruben F.

(ii) TITLE OF INVENTION: Sequences Characteristic of Human Gene Transcription Product

(iii) NUMBER OF SEQUENCES: 2412 (1-2421, with 9 SEQ ID NOS unused.)

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Knobbe, Martens, Olson, and Bear
(B) STREET: 620 Newport Center Dr. Sixteenth Floor
(C) CITY: Newport Beach
(D) STATE: CA
(E) COUNTRY: USA
(F) ZIP: 92660

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: 07/837,195
(B) FILING DATE: 12-FEB-1992

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: US 07/716,831
(B) FILING DATE: 20-JUN-1991

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Israelsen, Ned A.
(B) REGISTRATION NUMBER: 29,655
(C) REFERENCE/DOCKET NUMBER: NIH004.004CP1

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: 619-235-8550
(B) TELEFAX: 619-235-0176

SEQ ID NO:1: (Length of Sequence = 362 Nucleotides)

CTTCCCTTTT GTTCCCCTCA GTGTCCCTTT TAATTGCTTC CCTCCATTTT CCTTAGCAGC ATCCTAGTTG ATGGTCTGGG
TTATCAGAGG AGCAAAAACA TTTAAGTGTC AAATAATGCT CATTGTCTCC CTGGGATTTC TAAACAGAAA AAATGAAGAA

CACTACGGCG GCGTGGTCAG CCTCATCCCC CTCATCCTAG ACTTAATGAA AGAATGGWTC GCCCACTTCG AGAAGTTGCC
GCGCAAGGTG CTGCAGGGCA CGCGGGCTGC CTGCACT

SEQ ID NO:1931: (Length of Sequence = 343 Nucleotides)
ATCACTTCCC CACCCCACAG GATCTGCCCC AGAGAAAGTC CTCGCTCGTC ACCAGCAAGC TTGCGGGTGG CCAAGTTGAA
TGATGCTGCC CGGGGCTCTG CCAGATCCTG AGACGCTTCC CCTCCCTGCC CCACCCGGGT CCTGTGCTGG NTCCTGCCCC
TTCCTGCTTT TGCAGCCAGG GGTCAGGAGG TGGCTCGGGT GTGGGCTGGA GAGGCAGAAG CCCTTTCCTG TTGGTGTCCC
AGCACATGGA GCCCCTTGGG CTGAGCACCA AGACCTTGAA CCTTTTTTGT TTTACCTTTT TTCCAAATAA CAGTTTGGAG
AAATATCAAT GAAATCTGGG GGT

SEQ ID NO:1932: (Length of Sequence = 314 Nucleotides)
TTTCATGGGT TTTTGTTTTG TTTATTTCGA ATACTGAAAA AGTCCTTTGG GCTCTGTGGG GTTCCCCACG CTCACGGCTC
CTTTCTCCCA CACTCACTGC CCTTCTTCCC ACAGCAAATC TATTTCAAGG ACAGTACTTT TTAAAATGAT TAATGTTGAG
TTCTCAACTA GCTCTGCAGA ACTAGAGGAG CTGTTTGCAT CTGTCTGTGC GGATGGAGTT TCTTTTATCT GACACCAGGT
CTCCAACCAC ACTGATGCAA GGCATTTTAT CTACAGAGCT CAACTAGAAC CCCTTTTTCA TTAGGCTACT CCAA

SEQ ID NO:1933: (Length of Sequence = 378 Nucleotides)
AGCTTCCTGC GGGACCACAG CTATGTGACT GAAGCTGACA TCATCTCTAC CGTTGAGTTC AACCACACGG GAGAGCTGCT
GGCCACAGGT GACAAGGGCG GCCGGGTCGT CATCTTCCAG CGGGAACCAG AGAGTAAAAA TGCGCCCCAC AGCCAGGGCG
AATACGACGT GTACAGCACT TTCCAGAGCC ACGAGCCGGA GTTTGACTAT CTCAAGAGCC TGGAGATAGA GGAGAAGATC
AACAAGATCA AGTGGCTCCC ACAGCAGAAC GCCGCCCACT CACTCCTGTT CCACCAACGA TAAAACTATC AAATTATGGA
AGATTACCGA ACGAGATAAA AGGCCCGAAG GATACAACCT GAAGGATGAA GAGGGGAA

SEQ ID NO:1934: (Length of Sequence = 239 Nucleotides)
ATTTAAATTG ACAGCCTTCC ATTTTTCGAG AAAGTACAAA CAGAACTGCT TTAGCACCCA TCGAGCCCCA AACGGGTAAG
GTAAGCCAAG GTTTTAATGA CCAGCCCAGT ATCTAAGCTT CCAAACGGAT GCCAGCCCAT CACATACTYA CCCTGGGAGG
CTGCTGCACG GGCATTCTCC YGATGCTCAC GGCACTTGGK GTAGGTTTCA RGATCGCCTC TTTGAGGAAG GACTTCAGG

SEQ ID NO:1935: (Length of Sequence = 319 Nucleotides)
TTAATTTTTT TNTCCCATAG AGGAATAGCA TTACAGTCTA ACAATCAGAA TTCTGTTACA CACATACACA GGCATGCCAC
ATGACCCAGT TGAGGTGGTT GTNTCCTTGA GTCTGTTGAC ACGTCACATG GTCAAAGTCT CCTCATTTCA GCCAGTCTCA
ACACAAAACA CCCAACAGGG ATGCACTCAA CTTGTTGGTT CCATGTGGAA CTAGGTGGCA GGGCGAGAGG GAAAGTAGTA
GAAGGGGGCT ATGGTGTGTC TGCATTCAGT CCCCTCACAT AAAGCCACAT GGATCTAGGG GGGGTATCCA AGAGCTCTG

SEQ ID NO:1936: (Length of Sequence = 415 Nucleotides)
CTATTTTTAC AAATTATACC TAATGAGTAA AATTAGTGTA AAGTGATAAC ATGCTTCTAC CTGTATTTCT AGTGACCCTT
TAGCGGCAGG TATTTATACC TGGTATTTAT GATGCAGTAT ATAAGTGGTG AACAATAACT GACAGTATTG TGCTTGCTGT
ACATGTCTGG TCTTTTGAAA CAGATTTTAG TAAGCATTTT CCAGAGGTAA AACTGTGTCC TTATTCTAAT TTTATTCCTA
GGGCAAAGTA GACAGGGATT ATTTCCTTGA ATCTATTTCC AAATTAATAT TTTTTTCTTT GGTATTTCTA CACTTTAAGG
CCATTTGGTG CAATTTAGAA AGTGTTGGCC TCCCTTCCGC TAGCCACATT CAAAATTAAC TTCCAAAACC TCAGGAACAG
TACAAGGAAT TTGAA